Copyright
by
Gezheng Wen
2017

The Dissertation Committee for Gezheng Wen
certifies that this is the approved version of the following dissertation:

Model Observer for Optimizing Digital Breast Tomosynthesis for Detection of Multifocal and Multicentric Breast Cancer

Committee:

Mia K. Markey, Supervisor

Alan C. Bovik

Joydeep Ghosh

Tamara Miner Haygood

Subok Park

James W. Tunnell

Model Observer for Optimizing Digital Breast Tomosynthesis for Detection of Multifocal and Multicentric Breast Cancer

by

Gezheng Wen, B.E.; M.S.E

DISSERTATION
Presented to the Faculty of the Graduate School of
The University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN
May 2017

Dedicated to my parents.

Acknowledgments

I would like to first extend my deepest gratitude to my parents Jinsheng Wen and Shuxian Yang. They have been a constant source of strength, inspiration, and emotional support for me. It is impossible to describe in words or numbers how much effort and sacrifice my parents have made through my journey, especially during tough periods of my life. I would have certainly not achieved this much without their consistent support and encouragement.

I am extremely fortunate and sincerely thankful to have Dr. Mia K. Markey as my Ph.D. advisor. Mia has not only guided me with her remarkable academic advice, but also provided invaluable support in my personal life and professional developments. She has always been extremely patient, considerate, and understanding through these years. I am especially thankful to Mia's continuous support and advice when I was away from my research for internships. I have benefited so much from what I have learned from Mia, such as her excellent time management skills, her attention to details, and her cheerful spirit. Moreover, I would like to sincerely thank Dr. Subok Park and Dr. Tamara Haygood, whose roles have been as important as my co-advisors. With her knowledge and experience in the research field of this dissertation, Subok's guidance and feedback have significantly helped in defining and de-

veloping the work presented in this dissertation. Her constructive advice has always inspired me to think deeper and more comprehensively, which has contributed to my growth as a researcher. I also appreciate the opportunity to collaborate with Tamara on a variety of research projects since the first year of my doctoral study. She has always been extraordinarily supportive and encouraging for my research. She has greatly helped me out with her clinical expertise and resources that have been critical in completing the work. I also acknowledge my committee members, Dr. Alan C. Bovik, Dr. Joydeep Ghosh and Dr. James W. Tunnell, for their helpful feedback on this dissertation.

I would also like to thank current and former BMIL members, particularly, Dr. Gautam Muralidhar, Dr. Nishant Verma, Dr. Juhun Lee, Dr. Clement Sun, Nisha Kumaraswamy, Krista Nicklaus and Yao Zhang. I am especially thankful to Gautam and Nishant, who have been like mentors to me since the day I joined the lab. I would also like to thank the staff of Biomedical Engineering, Electrical and Computer Engineering, and Information Technology Services. In particular, Michael Don, Melanie Gulick and Carter Francis have helped me through numerous administrative and technical issues. Moreover, I would like to express my sincere thanks to the scientists and staff in the Division of Imaging, Diagnostics, and Software Reliability at FDA, particularly, Dr. Rongping Zeng, Dr. Jonathan Boswell and To-Oanh Pham, for their support on this dissertation. Finally, I would also like to thank the National Science Foundation for providing the financial support to my research at FDA.

Model Observer for Optimizing Digital Breast Tomosynthesis for Detection of Multifocal and Multicentric Breast Cancer

Publication No. _______________

Gezheng Wen, Ph.D.
The University of Texas at Austin, 2017

Supervisor: Mia K. Markey

The goal of medical imaging is to acquire and display images of human anatomy and function such that they can be optimally interpreted by a trained observer, e.g., a radiologist. Start-of-art medical image quality is measured by the performance of an observer on a given clinical task. Since psychophysical studies are resource intensive, model observers are widely used as a surrogate in task-based assessment of image quality. Model observers are typically designed to detect at most one abnormality, e.g., a single lesion. However, in clinical practice, there may be multiple abnormalities in a single set of images, which can have a significant impact on treatment planning and outcomes. For example, patients with multifocal and multicentric breast cancer (MFMC), i.e., the presence of two or more tumor foci within the same breast, are more likely to undergo mastectomy rather than breast conservation therapy. Detecting multiple breast tumors is challenging because the prevalence of tumors varies significantly across breast regions, and radiologists do not know the number or location of tumors a *priori*.

The vision of this dissertation is that digital breast tomosynthesis (DBT) has the potential to improve the detection of MFMC, and may offer advantages such as fewer false-positive findings, lower cost, and better accessibility. This dissertation focuses on the design and applications of a model observer to optimize DBT system geometries for detection of multiple breast tumors. This is significant and innovative because prior efforts to optimize DBT image quality only considered unifocal breast cancer scenarios. We highlight the following two main aspects of contributions in this dissertation: (1) We have developed a novel model observer that detects multiple abnormalities in anatomical backgrounds. (2) We have employed the extended 3D multi-lesion model observer to identify DBT system geometries that are most effective for the detection of MFMC. Our results demonstrate that the presence of more than one tumor present distinct challenges to DBT optimization, and that DBT geometries that yield images that are informative for the task of detecting unifocal breast cancer may not necessarily be informative for the task of detecting MFMC. We are validating the clinical relevance of our model observer studies with an ongoing human observer study with experienced breast imaging radiologists.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background & Significance

The objective of medical imaging research is to create imaging systems and image presentation methods to achieve accurate, safe, timely, and cost-effective diagnosis and treatment of disease. Accurate interpretation of medical images is imperative for successful patient treatment. Hence, the state-of-the-art definition of medical image quality is task-based: how well can the information necessary for a given task of clinical interest be extracted from the images? [61]. Specifically, medical image quality is measured by the average performance of multiple observers in conducting the task.

In the development and evaluation of imaging systems, there are many modality-specific parameters (e.g., projection angular increments in digital breast tomosynthesis (DBT)) that impact the observers ability to extract diagnostic information from the image. When designing new imaging systems, these parameters must be adjusted to optimize observer performance. Likewise, when there are proposed changes to existing image acquisition, processing, and display methods, an assessment regarding the efficacy of the potential innovations is desired. Both scenarios require a task-based assessment of the image quality. However, it is expensive in terms of both money and effort to conduct such studies with human observers (e.g., radiologists), especially when the goal is to study the effects of multiple variables on image quality. Model observers are widely used as surrogates in task-based assessments and optimizations of medical image quality (e.g., [119, 78, 160, 102, 104, 101]). A typical example of using model observers in medical imaging research is given by Reiser et al. [119], in which the authors used a model observer to examine the effects of scan parameters on lesion detectability on DBT (e.g., with a fixed angular span, detection performance with different numbers of projection views was ranked). Moreover, well-designed simulation studies are essential complements to bench testing and clinical studies because simulations are flexible and cost-efficient for evaluating many system configurations (e.g.,[119, 56, 158]). This can help narrow down parameter spaces that require further evaluation. Similar studies are usually difficult if not infeasible to conduct with physical phantoms or patients since those designs are limited to a modest number of images. In addition, physical phantoms can lack realism or variation, and some studies could put patients at risk (e.g., x-ray exposure).

Most existing model observers are designed to detect at most one abnormality [61]. Previous model observers for detecting multiple signals rely heavily on unrealistic assumptions. For example, Popescu [115] derived the signal specificity function from likelihood ratio tests for detecting multiple signals, but under the assumption that the scores for the signals and noise features were independent and stationary. Banerjee et al. [7] used a visual search human-model observer for a multitarget tumor localization task, but the channelized non-pre-whitening discriminant in the analysis stage of the model assumed that background was known exactly. Such assumptions are rarely satisfied in real medical images. For example, studies (e.g., [110, 87, 120]) have shown that local breast density differs significantly across breast subregions, and the distribution of radiodense breast tissue is spatially autocorrelated. Thus, it is critical for model observers to account for these factors. This would facilitate assessment and optimization of image quality for improving human readers' interpretation of cases with potential multiple abnormalities.

Multifocal (MF) breast cancer is commonly defined as two or more tumor foci within a single breast quadrant. Similarly, multicentric (MC) breast cancer is defined as two or more tumor foci within different quadrants of the same breast. The prevalence of MF and MC breast cancers (MFMC) has been reported to range from 9% to 75%, due to a lack of consistent assessment methods [167, 95, 39]. In the largest study of 8,935 breast cancer patients [155], the disease incidence was 20.8% (79.2% of the patients had a unifocal disease, 15.6% had MF tumors, and 5.2% had MC tumors). MFMC can result from intra-mammary spread from a single primary tumor, or multiple synchronous primary tumors with inter-tumoral heterogeneity [41, 3, 91]. A diagnosis of MFMC significantly impacts treatment planning. For instance, patients with MFMC are more likely to undergo mastectomy (i.e., the whole breast removed) than breast-conserving therapy (i.e., only the tumor and some surrounding tissue removed). Though the impact of MFMC on prognosis is still an ongoing controversy, MFMC is generally more aggressive than unifocal breast cancer (e.g., higher recurrence and metastatic rate [167, 95, 155]). It is desirable to locate all of the lesions in the breast prior to the commencement of treatment as the objective is to remove all of them at the same time.

Detecting multiple breast tumors is challenging as MFMC breast cancers are relatively uncommon, and the number and location of tumors isn't known a priori. The prevalence of tumors also varies across breast regions (e.g., almost half of breast cancers are detected in the upper outer quadrant of the breast [80]). It is usually the case that the more conspicuous lesion is found first, and so finding the second lesion is more difficult as it will probably be smaller or in other ways less obvious. Cytology studies (e.g., [20, 43, 88]) have shown that the prevalence, and conspicuity of a particular type of targets may lead to missing subtle, smaller targets. Fleck et al. [53] showed in a multi-target search study that readers missed low-contrast targets more often when a high-contrast target was detected than on trials in which the low-contrast target was presented alone. Additional factors, such as bias about the expected number of tumors [53], satisfaction of search (i.e., detection of one target makes detection of subsequent targets less likely)[13, 125, 12], may also impact the process of finding the second tumor. Women with breast cancers often undergo magnetic resonance imaging (MRI) to check for the presence of additional foci of disease that may be occult on mammography. While MRI is very sensitive for detecting MFMC, it has low specificity ($37\% - 86\%$ [82]). Thus, pre-operative MRI can lead to additional breast imaging, additional breast biopsies, and/or overtreatment, e.g., false-positive findings can result in women undergoing mastectomy rather than breast conserving therapy.

DBT may improve the diagnosis of MFMC. DBT is a new 3D mammography, first approved by US Food and Drug Administration in 2011, formed by sweeping an x-ray beam in an arc over a limited angular range [127]. A series of low-dose projection images (usually 9-30) are acquired at different angles, and then reconstructed into a stack of 2D slices by shifting and adding the individual projections [127, 128]. DBT reduces tissue overlap relative to conventional mammography, and thus has the potential to improve breast cancer detection (i.e., higher sensitivity), decrease false alarms (i.e., higher specificity), and more accurately characterize lesions (e.g., [109, 59, 141, 22, 97]). A few clinical studies have suggested that DBT has the potential to improve the detection of MFMC relative to MRI. For example, Wasserman et al. reported a case in which multiple foci were occult on MRI, but correctly identified on DBT [144]. These studies suggest that the potential sensitivity of DBT for MFMC may be equivalent to or higher than that of MRI, although no large comparison study has been published to date. DBT is expected to have a significantly higher specificity than MRI because the major cause of false positives on MRI (e.g., non-mass enhancement caused by factors such as mastopathic changes, fibrocystic changes due to hormonal stimulation [92]) is not relevant in x-ray imaging such as DBT. There are ongoing clinical trials (e.g., TOMODIAG in France [143]) that aim to investigate the potential of DBT in MFMC diagnosis. Overall, DBT may offer advantages over MRI in terms of fewer false-positive findings, lower cost, and better accessibility. In addition, DBT may be a better alternative than ultrasound when MRI is contraindicated; for example, Clauser et al. found a 23% increase in detecting additional lesions with second-look DBT relative to ultrasound [37]. However, prior efforts to optimize DBT system geometry for image quality only considered unifocal breast cancer scenarios (e.g., [127, 119, 66, 57, 130, 50]). DBT system geometries that yield images, that are informative for the task of detecting unifocal breast cancer, may not necessarily yield images that are ideally informative for the task of detecting MFMC. For example, a lesion seen in the specific section of reconstructed images is in focus, but the out-of-focus blurring of the lesion in the depth direction may affect the detectability of another lesion, depending on its size and its distance from the in-focus lesion (e.g., [109, 89]). Lee et al. [81] also showed that the detectability of small masses on DBT images varied across different breast regions in both in-plane and depth directions. Such factors have not been considered in prior optimization efforts with unifocal cases. Thus, optimizing DBT image quality for the detection of MFMC is important to enhance the potential clinical value of DBT.

Different DBT system designs can yield different image characteristics such as spatial resolution [66, 165], lesion detectability [119, 40, 32], contrast-to-noise ratios [84] and observers' preference [57]. There are two key design choices in image acquisition [127] that we are particularly interested in: 1) narrow- or wide-arc geometry (i.e., small or large angular span); and 2) a large or small increment of projection angles. An optimal DBT acquisition geometry can be obtained by balancing tradeoffs between and interaction effects of these two factors among many different geometry parameters. For instance, a larger angular span usually improves the depth perception of the breast, but it also decreases in-plane spatial resolution (e.g.,[119, 86]). On the other hand, using a larger number of projection angles improves angular sampling. However, at a fixed total x-ray exposure and a fixed angular span, it also increases the amount of quantum noise in the reconstructed images [130]. In addition, we are interested in the impact of distribution of projection views. Different projection-view distributions affect the degrees of in-plane and inter-plane blurring in opposite ways [84]. Increasing the angular span and the number of projection angles generally leads to a longer total scan time that may result in increased motion artifacts [57]. Moreover, considering multiple tumors for DBT quality assessment raises new questions. For example, as tumors in MF cases are physically closer to each other than in MC cases, would higher in-plane resolution lead to higher lesion detectability? Would the anatomical variations in local breast regions as in MC cases cause less stable detection performance? These questions remain to be answered.

1.2 Dissertation Contributions

This dissertation is divided into two parts. The first part of this dissertation is focused on the model observer design for detecting multiple abnormalities in a single set of images. As described earlier, existing model observers can typically detect at most one signal, and previous studies involving multiple signals rely heavily on unrealistic assumptions. Moreover, the background of medical images is usually non-stationary due to the presence of anatomical structures, and such variations across image regions may influence the observer performance. Hence, our first contribution is a novel model observer that overcomes these limitations, and addresses multiple-signal detection tasks [151, 150, 152]. Without prior knowledge of image backgrounds and signals, our multi-signal model observer not only incorporates interactions among signals, but also accounts for variations in local anatomical background by adjusting the decision-making rule accordingly. We develop and deploy novel channelization approaches to estimate efficient channels that capture significant discriminant image information, such as local background statistics, signal characteristics,and signal prevalence across locations. The evaluation studies with synthetic 2D mammograms demonstrate that the model observer is able to detect multiple breast lesions, making accurate image-level and location-specific detection decisions. This shows the potential of the model observer as a broadly applicable tool in medical image quality assessments of realistic clinical situations. This would greatly facilitate the task-based optimization of imaging systems.

The second part of this dissertation is focused on optimizing DBT image quality for the detection of MFMC breast cancer. The motivation is that with proper selection of system parameters, DBT may be a less expensive, more accessible, and more reliable alternative to breast MRI for detecting MFMC. However, as described earlier, previous DBT system optimization studies are based on unifocal breast cancer scenarios, and no study has sought to optimize DBT for detecting multiple breast tumors, or to investigate how the presence of more than one tumor would influence DBT system design. Our first contribution in this part of the dissertation is that we extend our 2D multi-signal model observer into a 3D model observer that works effectively with volumetric DBT imaging data [153, 149]. The proposed 3D multi-lesion (ml) channelized Hotelling observer (CHO) considers 3D DBT reconstructed image data as a single object to be examined, and makes detection decisions regarding the presence of lesions upon the overall information from both individual slices and their correlations. In the evaluation studies with simulated DBT images of MFMC cases (i.e., digital breast phantoms with multiple embedded synthetic breast lesions were scanned by a simulated DBT scanner), we demonstrate that the 3D ml-CHO could achieve high detection performance with a small number of channels. We also show that incorporating locally varying anatomical backgrounds and their correlations as in the design of 3D ml-CHO is beneficial for making multi-lesion detection decisions. Moreover, our second contribution is that we employ the 3D ml-CHO model observer to identify DBT system geometries that are most effective for the detection of MFMC [153, 148]. We have shown that the consideration of multiple breast lesions may present distinct challenges to DBT system optimization. We have demonstrated that DBT system geometries, especially wide-arc geometries versus narrow-arc geometries, may not be equally effective for detecting MF or MC lesions. We have also shown that the rank orders of the geometries for MF and MC cases and for unifocal cases are different. Thus, our model observer evaluation studies highlight that the optimal geometry of DBT may vary when the task of clinical interest changes, and a given DBT system geometry may not yield images that are equally informative for detecting MF, MC, and unifocal breast cancers. Our final contribution is an ongoing human observer study with experienced breast imaging radiologists [147], through which we will validate the clinical relevance and significance of our model observer studies. The radiologists perform the mimicking task of MFMC detection by interpreting simulated DBT images acquired with different imaging geometries. They score each image according to their confidence of the presence of a lesion at each of the potential lesion locations. It is hypothesized that the rank ordering of DBT systems by the human observer performance would generally agree with the performance trends shown in our model observer studies. Overall, all these conclusions are significant and constructive because they could potentially help enhance the clinical values of DBT, and improve cancer management of MFMC.

1.3 Dissertation Outline

The remainder of this dissertation is organized as follows. In Chapter 2, we briefly review existing work on model observers and their applications in medical imaging research. In Chapter 3, we present a novel model observer for detecting multiple abnormalities in anatomical background images. In Chapter 4, we present background materials on MFMC breast cancer, and the potential of DBT in improving the detection of MFMC. In Chapter 5, we present an extended 3D model observer that detects multiple breast lesions from volumetric DBT imaging data. Chapter 6 presents a model observer study in which we identify DBT system geometries that are most effective for the detection of MFMC. In Chapter 7, we present an ongoing human observer study with experienced breast imaging radiologists in which we will validate the clinical relevance of the model observer studies. In Chapter 8, we conclude this dissertation with pointers to interesting areas for future work.

Chapter 2

Model Observer in Medical Imaging Research

2.1 Overview

In this chapter, we briefly review the fundamentals of model observers, and their applications in medical imaging research (Section 2.1). One of the main objectives of this chapter is to emphasize the limitations of existing model observers due to their lack of abilities in detecting multiple abnormalities from a single image set (Section 2.3). Such drawbacks have motivated us to develop a practical multi-signal model observer in this dissertation. Section 2.2 reviews the key aspects of detection tasks in Section 2.2.1, CHO in Section 2.2.2 and 2.2.3, channelization approaches in Section 2.2.4 and extended CHOs for volumetric imaging data in Section 2.2.5. Moreover, we introduce a number of frequently used notations that will be used throughout this dissertation.

The objective of medical imaging research is to create imaging systems and image presentation methods to achieve accurate, safe, timely, and cost-effective diagnosis and treatment of disease. Accurate interpretation of medical images is imperative for successful patient treatment. Hence, the state-of-the-art definition of medical image quality is task-based: how well can the information necessary for a given task of clinical interest be extracted from the images? [61]. Specifically, medical image quality is measured by the average performance of multiple observers in conducting the task. Clinical tasks can be roughly grouped into three categories: 1) detection tasks, such as a task of finding a tumor in the image, 2) estimation tasks, such as a task of estimating the size and location of a tumor, or 3) classification tasks, such as a task of classifying detected tumors into benign or malignant. As an invaluable component of medical image science, task-based image quality assessment is based on the fundamental idea of using mathematical and statistical theories to represent the clinical tasks as equations to be solved [11, 9, 10]. There are four key elements that must be carefully considered when conducting such task-based assessment in research or applications: a task of clinical interest, a specification of the population, an appropriate observer, and a figure-of-merit (FOM) [61]. For instance, the quality of screen mammograms could be represented by the performance of an observer, either a real radiologist or a mathematical model, in the well-controlled diagnostic task of making a binary detection decision regarding the presence of any sign of breast cancer. An ensemble of mammograms from normal and abnormal patients serves as the population. The area under a receiver operating characteristics (ROC) curve is estimated as the FOM.

In the development and evaluation of medical imaging systems, there are many modality-specific parameters (e.g., projection angular increments in

DBT) that impact the observer's ability to extract diagnostic information from the image. When designing new imaging systems, these parameters must be adjusted to optimize observer performance. Likewise, when there are proposed changes to existing image acquisition, processing, and display methods, an assessment regarding the efficacy of the potential innovations is desired. Both scenarios require a task-based assessment of the image quality. However, it is expensive in terms of both money and effort to conduct such studies with human observers (e.g., radiologists), especially when the goal is to study the effects of multiple variables on image quality. Model observers are widely used as surrogates in task-based assessments and optimizations of medical image quality (e.g., [119, 78, 160, 102, 104, 101]). A typical example in imaging research practice is given by Reiser et al. [119], in which the authors used a model observer to examine the effects of scan parameters on lesion detectability on DBT (e.g., with a fixed angular span, detection performance with different numbers of projection views was ranked). There are a large number of such applications of model observers in both hardware and software system optimization. Different designs of model observer may be desirable for these two general purposes. For hardware optimization (e.g., investigating how different system parameters affect tumor detection in the presence of noise), a model observer that extracts as much statistical information as possible from the images is needed (e.g., [2, 58]); while for software system optimization (e.g., comparing image rendering algorithms), a model observer that predicts human observer performance when given the task of interest in a consistent manner is needed (e.g., [133, 142, 33]).

2.2 Channelized Hotellling Observer

2.2.1 Binary Detection Task

In a binary detection task, there are two possible hypotheses for a given image: signal present (e.g. tumor present) or signal absent (e.g. normal tissue). Mathematically, if we denote $\mathbf{s}$ as the signal to be detected, $\mathbf{b}$ as the noise-less image background, and $\mathbf{n}$ as the measurement of noise in the image, the image data $\mathbf{g}$ under the signal-present H_1 and signal-absent H_0 hypotheses are given by:

$$\begin{aligned} \mathrm{H}_0 : \mathbf{g} &= \mathbf{b} + \mathbf{n}, \\ \mathrm{H}_1 : \mathbf{g} &= \mathbf{b} + \mathbf{n} + \mathbf{s}, \end{aligned}$$

A task is called a signal-known-exactly task if the signal $\mathbf{s}$ is known exactly. Similarly, there are signal-known-statistically, background-known-exactly and background-known-statistically tasks.

2.2.2 Hotelling Observer

A linear observer applies a linear template $\mathbf{w}$ to an image data vector $\mathbf{g}$ to obtain the value of a scalar test statistic t:

$$t = \mathbf{w}'\mathbf{g}$$

The values of the test statistic t determines whether the hypothesis H_0 or H_1

are more likely to be true: the larger t is, the higher the probability of signal-present. The signal-to-noise ratio (SNR) is often used as a FOM to assess the ability of a linear observer in differentiate between classes of images [25], where SNR is defined as:

$$\mathrm{SNR} = \frac{< S >_{\mathbf{H}_1} - < S >_{\mathbf{H}_0}}{\sqrt{\frac{1}{2}(\sigma^2_{\mathbf{H}_0} + \sigma^2_{\mathbf{H}_1})}}$$

where $< S >_{\mathbf{H}_i}$ $(i = 0, 1)$ is the mean of the decision variable under $\mathbf{H}_1$ or $\mathbf{H}_0$, and $\sigma^2_{\mathbf{H}_i}(i = 0, 1)$ is the variance of the decision variable under $\mathbf{H}_i$.

The Hotelling observer (HO) is the optimal linear observer in terms of maximizing SNR [61]. The basic idea is to use the mean and covariance of image data in its decision strategy to classifying the two classes. Its decision template $\mathbf{w}_{\mathrm{HO}}$ is computed:

$$\mathbf{w}_{\mathrm{HO}} = \mathbf{K}_{\mathbf{g}}^{-1}\overline{\mathbf{g}_{\mathbf{s}}}$$

where $\overline{\mathbf{g}_{\mathbf{s}}}$ is the mean difference signal, and $\mathbf{K}_{\mathbf{g}}$ the average covariance matrix of signal-present and signal-absent images [25, 114]. It is a common practice to use separate datasets to train and test the HO.

2.2.3 Channelized Hotelling Observer

Due to the high dimensionality of image data, the computational burden of calculating the covariance matrix and its inverse is extremely high. Channelization is one of the most widely used approaches to solve this problem, and the resulting model observer is called channelized Hotellling observer

(CHO) [61]. A channelized image $\mathbf{U}$ can be represented as:

$$\mathbf{U} = \mathbf{T}\mathbf{g}$$

where $\mathbf{g}$ is the image with or without the signal, $\mathbf{T}$ is an $N_c \times N_p$ matrix that represents a set of carefully designed channels for extracting relevant image information, and $\mathbf{U}$ is an $N_c \times 1$ channelized image. N_c is the number of channels used to reduce the dimension of the image data and N_p is the original dimension of the image data. Generally, N_c is much less, even several orders of magnitude less than N_p. The scalar test statistic t of the linear discriminant is in the form of:

$$t(\mathbf{U}) = \mathbf{w}_{\mathbf{U}}^T\mathbf{U}$$

where $\mathbf{w}_{\mathbf{U}}$ is the data space template in the form of:

$$\mathbf{w}_{\mathbf{U}} = \mathbf{K}_{\mathbf{U}}^{-1}\overline{\mathbf{U}_{\mathbf{s}}},$$
$$\overline{\mathbf{U}_{\mathbf{s}}} = \overline{\mathbf{U}_1} - \overline{\mathbf{U}_0}, \quad \overline{\mathbf{U}_j} = \mathbb{E}[\mathbf{U}|H_j], j = 0, 1$$
$$\mathbf{K}_{\mathbf{U}} = \tfrac{1}{2}[\mathbf{K}_{\mathbf{U},0} + \mathbf{K}_{\mathbf{U},1}], \mathbf{K}_{\mathbf{U},j} = \mathbb{E}[(\mathbf{U} - \overline{\mathbf{U}_j})(\mathbf{U} - \overline{\mathbf{U}_j})^T|H_j], j = 0, 1$$

Here, $\overline{\mathbf{U}_{\mathbf{s}}}$ is the mean difference channelized signal, and $\mathbf{K}_{\mathbf{U}}$ the average covariance matrix of signal-present and signal-absent channelized images [25, 114].

There are two main categories of channels: anthropomorphic and efficient, where anthropomorphic channels are derived from the characteristics of human visual system, and efficient channels are called efficient when the channelized observer (e.g., CHO) approximates the performance of the corresponding unconstrained observer (e.g., HO) [61].

(i). **Anthropomorphic channels**: There are a large number of anthropomorphic channel designs (e.g., difference-of-Gaussians channels [1], Gabor channels [162]) that have been shown to be highly correlated with human observer performance. Spatial-frequency selective mechanism is one of the most commonly modeled characteristics, as human visual system is selectively sensitive to various ranges of spatial frequencies [123].

(ii). **Efficient channels**: Efficient channels are mainly used to estimate the unconstrained observer performance while reducing the image data dimension with as few channels as possible [61]. Factors such as physics of the imaging system, signal and background data statistics should be incorporated when designing a set of efficient channels for the given task. Practically, a finite number of basis functions are often chosen from a complete set of basis functions to approximately describe the high-dimensional matrices (e.g., an image ensemble [154], an HO template [54]). This could help remove redundant information in the images, especially when the information is not relevant to the task. Some of the widely used efficient channel design include Laguerre-Gauss (LG) channels [54], partial least squares (PLS) channels [154], and singular vectors (SV) channels [106].

2.2.4 Choosing Efficient Channels

In this dissertation, we are particularly interested in efficient channels that consider signal and background data statistics in the task. Among possible choices for channelization, we describe two sets of the most promising efficient channels (i.e.,LG channels and PLS channels) as well as their advantages and disadvantages.

(i). **Laguerre-Gauss channels** [54]: LG channels,as the product of Laguerre polynomials and Gaussian function, are often used in task-based assessments of medical image quality (e.g., [32, 158]). One key advantage of LG channels is that LG channels are based on an analytical expression, so the quality of LG channels is independent of the number of available training samples [61]. To make LG channels efficient for the CHO, it is important to adjust the parameters of LG channels, including the Gaussian width and the number of LG channels to be used. Prior studies have shown that the Gaussian width need to be large enough to cover the speared of the signal [54, 103, 160] and the optimal number of LG channels varies as the background and signal statistics changes [61, 154]. However, LG channels only work well with rotationally symmetric signals (e.g., circular Gaussian functions) because they naturally formulate an orthornamal basis for rotationally symmetric square-integrable functions (i.e., the correlation structure of the background has no preferred orientation) [61].

(ii). **Partial least squares channels** [154]: PLS channels have been shown to effectively extract relevant information for signal detection tasks and approach the performance of an ideal observer (e.g., [161, 107, 101]). The idea of the PLS approach is to seek a small set of channels that maximize the covariance between the image data and the decisions/truth labels [154]. As PLS channels are estimated directly from the images (without prior knowledge of the background or the signals), they are adaptive to complex image backgrounds and signal statistics. This is one of the main advantages of PLS channels over other choices of efficient channels. However, the quality of PLS channels highly depends on the number of available channel training images [154].

2.2.5 CHO for Volumetric Imaging Data

Given the increasing use of consecutive cross-sectional medical imaging modalities in clinical practice (e.g., computed tomography (CT), DBT), advanced multi-dimensional CHO models have been recently developed, and applied to imaging research. For example, Platisa et al. [114] compared three designs of multi-slice CHO, and a volumetric CHO for the task of detecting 3D signals in volumetric images. Park et al. [104] developed a 3D-projection HO that incorporated spatial correlation between the angular projections of a DBT scan through concatenating different angular projections. Kim et al. [72] showed that a volumetric CHO using 3D channels significantly outperformed 2D CHO observers in detecting tumors from the whole-body PET imaging.

Chen et al. [35] proposed a volumetric CHO design that first uses a 2D CHO for each slice and each view of simulated SPECT myocardial images, then integrates all the resulting test statistics for the final test statistics. The key difference among these extensions of CHO models is how the models integrate information from multiple slices of the volumetric imaging data, and how the overall information is used to make a detection decision.

2.3 Lack of Multi-Signal Model Observer

Most existing model observers are designed to detect at most one abnormality [61]. However, in clinical practice, there may be multiple abnormalities in a single set of images (e.g., MFMC breast cancer, multiple pulmonary nodules in chest CT scans), and the presence of more than one abnormality can have a significant clinical impact on treatment planning. For instance, patients with MFMC breast cancer are more likely to undergo mastectomy rather than breast conservation therapy [95]. Detecting multiple abnormalities is challenging as the prevalence of abnormalities can vary substantially across image regions (e.g., almost half of breast cancers are detected in the upper outer quadrant of the breast [80]), and human observers do not know the number or location of abnormalities a priori. New imaging techniques, such as DBT, have the potential to improve multiple-signal detection. For example, Wasserman et al. [144] reported a case in which multiple foci carcinomas were occult on breast MRI, but correctly identified on DBT. Thus, it is essential to develop a practical model observer for detecting multiple abnormalities in the same image set.

Previous model observers for detecting multiple signals rely heavily on unrealistic assumptions. For example, Popescu [115] derived the signal specificity function from likelihood ratio tests for detecting multiple signals, but under the assumption that the scores for the signals and noise features were independent and stationary. Banerjee et al. [7] used a visual search human-model observer for a multitarget tumor localization task, but the channelized non-pre-whitening discriminant in the analysis stage of the model assumed that background was known exactly. Such assumptions are rarely satisfied in real medical images. For example, studies (e.g., [110, 87, 120]) have shown that local breast density differs significantly across breast subregions, and the distribution of radiodense breast tissue is spatially autocorrelated. Thus, it is critical for model observers to account for these factors. This would facilitate assessment and optimization of image quality for improving human readers' interpretation of cases with potential multiple abnormalities.

In the next chapter, we present a novel model observer for detecting multiple abnormalities in anatomical background images. Our multi-signal model observer is designed to overcome the aforementioned limitations of existing model observers. Simulation studies with synthetic mammograms are present to demonstrate the potential values of the multi-signal model observer in medical imaging research.

Chapter 3

Multi-signal Model Observer

3.1 Introduction

In this chapter, we present a novel model observer for detecting multiple abnormalities in anatomical background images [1,2,3]. The model observer is designed to perform a multiple-signal detection task, making both image-level decisions (i.e., at-least-one signal-present versus signal-absent) and location-specific decisions (i.e., how many signals are present at the given possible signal locations). Novel implementations of PLS and LG have been developed to estimate different sets of efficient channels that potentially help approximate the performance of ideal observers. The PLS channels capture interactions

---

[1] **Gezheng Wen**, Mia K. Markey, and Subok Park. Model observer design for detecting multiple abnormalities in anatomical background images. *SPIE Medical Imaging*, **9787**: 97870S, 2016. (Gezheng Wen is the primary author who developed the methods, performed the analysis, and prepared the manuscript. Mia K. Markey and Subok Park helped with the study design and manuscript revisions.)

[2] **Gezheng Wen**, Mia K. Markey, and Subok Park. Influence of local anatomical variations on detection of multifocal and multicentric breast cancer. *Medical Physics*, **43**(6): 3817, 2016. (Gezheng Wen is the primary author who developed the methods, performed the analysis, and prepared the manuscript. Mia K. Markey and Subok Park helped with the study design and manuscript revisions.)

[3] **Gezheng Wen**, Mia K. Markey, and Subok Park. Model observer design for multi-signal detection in the presence of anatomical noise. *Physics in Medicine and Biology*, **62**(4): 1396, 2017. (Gezheng Wen is the primary author who developed the methods, performed the analysis, and prepared the manuscript. Mia K. Markey and Subok Park helped with the study design and manuscript revisions.)

between signals and the background that provide discriminant image information. Linear decision templates are employed for generating both image-level and location-specific decision variables regarding the presence of signals. Simulation studies with synthetic 2D mammograms and two sets of lesions were conducted to evaluate the model observer in terms of the accuracy of making detection decisions, efficiency in the number of channels needed, and robustness with respect to variations in image backgrounds. In the preliminary versions of this study presented and published in the proceedings of *SPIE Medical Imaging 2016* [151], and *AAPM Annual Meeting 2016* [150], we presented the initial design of the model observers, and an evaluation study with synthetic MF cases and circular Gaussian signals. In the extended journal manuscript published in *Physics in Medicine and Biology* [152] in 2017, we further explored the generality of the model observers, focusing on synthetic MC cases in which local background statistics are different across possible signal locations. The evaluation results highlight performance with realistically shaped signals. In all these three works, Gezheng Wen developed the methods, performed the analysis, and prepared the manuscript. Mia K. Markey and Subok Park helped with the study designs and manuscript revisions.

This chapter is organized as follows: Section 3.2.1 briefly summarizes the multi-signal detection task to be performed by the model observer; Section 3.2.2 presents the simulated image dataset; Section 3.2.3 to Section 3.2.6 describes the model observer design, and Section 3.2.7 covers the FOMs for evaluating the model observer.

3.2 Materials and Methods

3.2.1 Multi-signal Detection Task

There were three key assumptions on images and signals: 1) One or more signals may be present in an image; 2) There were a finite/small number of possible, pre-defined signal locations; 3) The prevalence of signals varied across the signal locations. The number of signal locations in Assumption 2 is carefully chosen to reflect the average number of localized areas that human observers may pay more attention to during visual search of the images (e.g., [76, 75, 77, 62]). For example, Nodine et al. [96] showed that during holistic search of a mammogram, there were four regions of attention on which radiologists spent more than 1000 milliseconds of dwell time. These regions overlapped with the suspicious locations, from which the positive findings of breast cancer were reported.

The task of the model observer is to perform a multiple-signal detection task and make:

(i). Image-level decision: An image is called signal-present if there is at least one signal whereas an image is called signal-absent if there is no signal.

$$\begin{aligned} \mathbf{H}_0 : \mathbf{g} &= \mathbf{b} + \mathbf{n} \\ \mathbf{H}_1 : \mathbf{g} &= \mathbf{b} + \mathbf{n} + \mathbf{s} \end{aligned}$$

Table 3.1: Notation of image data and response

Notation	**Dimension**	**Comments**
N	scalar	Number of images
N_p	scalar	Number of pixels in each image
N_l	scalar	Number of possible signal locations
$\mathbf{X}$	$N_p \times N$	Image data
$\mathbf{Y}^{\mathrm{L}}$	$N_l \times N$	Location-specific truth labels
$\mathbf{Y}^{\mathrm{I}}$	$1 \times N$	Image-level truth labels

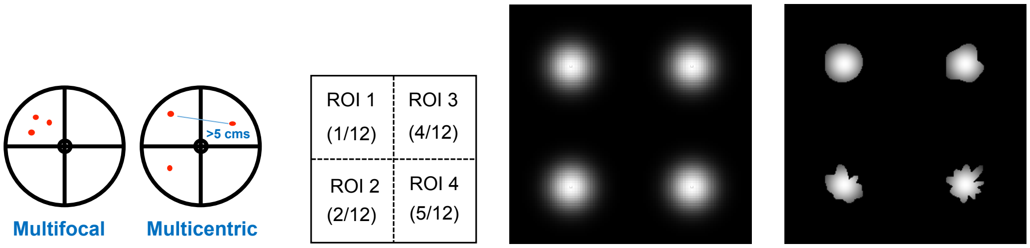
where $\mathbf{g}$ is image data, $\mathbf{b}$ is image background, $\mathbf{n}$ is the noise, $\mathbf{s}$ is the signal(s) to be detected, and $\mathbf{H}_i (i = 0, 1)$ are signal-absent and signal-present hypotheses, respectively.

(ii). Location-specific decision: For each possible signal location, a binary decision on signal presence is made.

$$\begin{aligned} \mathbf{H}_0(j) : \mathbf{g} &= \mathbf{b} + \mathbf{n} \\ \mathbf{H}_1(j) : \mathbf{g} &= \mathbf{b} + \mathbf{n} + \mathbf{s}_j, j = 1, 2, ...N_l \end{aligned}$$

where $\mathbf{s}_j (j = 1, 2, , N_l)$ is the signal to be detected at the j^{th} signal location, and $\mathbf{H}_i(j)(i = 0, 1)$ are signal-absent and signal-present hypotheses, respectively.

Table 3.1 summarizes the notation used in the study. The image dataset has a total of N images, and each image has a total of N_p pixels. Each column of the image data matrix $\mathbf{X}$ represents one image, and each column of response $\mathbf{Y}^{\mathrm{L}}$ is the location-specific truth labels of the image. $\mathbf{Y}^{\mathrm{I}}$ is the image-level truth label, and it is derived from $\mathbf{Y}^{\mathrm{L}}$ (i.e., $\mathbf{Y}^{\mathrm{I}}$ is 1 if at least one element of $\mathbf{Y}^{\mathrm{L}}$ in the same column is 1; otherwise, $\mathbf{Y}^{\mathrm{I}}$ is 0),



(a) (b) (c) (d)

Figure 3.1: (a) Illustration of MFMC breast cancer (red dots indicating the tumors), and the breast is divided into 4 quadrants. (b) Illustration of signal locations and their corresponding prevalence. (c) Signal-only image shows the four pre-defined signal locations with circular Gaussian signals. (d) Signal-only images with four irregular shaped signals that mimic real lesions. The degree of irregularity is 3, 5, 8 and 10, respectively.

3.2.2 Image Dataset

3.2.2.1 Image background

Clustered lumpy backgrounds (CLB) are commonly used in model observers studies to generate synthetic mammograms (e.g., [16, 28, 5]). By summing over sets of randomly positioned lumps centered around randomly positioned clusters [16], we simulated a dataset of CLB using MATLAB (MathWorks, Natick, MA). We assumed that N_l was 4, which mimics the fact that radiologists usually report the location of a detected lesion using the breast quadrant notation (i.e., lower inner quadrant, lower outer quadrant, upper inner quadrant, upper outer quadrant) [131]. The signal locations were sampled from a prevalence distribution: {ROI1(lower inner quadrant), ROI2(lower outer quadrant), ROI3(upper inner quadrant), ROI4(upper outer quadrant)} $= \{\frac{1}{12}, \frac{2}{12}, \frac{4}{12}, \frac{5}{12}\}$ (Figure 3.1(b)). The distribution was chosen to mimic the typical incidence of breast lesions detected in different breast regions (e.g., [80, 100, 15]). For example, primary tumor sites of 305,443 patients by breast quadrant was $\{8.7\%, 13.7\%, 22.6\%, 55\%\}$ [8].

(i). For multifocality, only one breast quadrant is involved, and the distance between tumors is usually less than or equal to 5 cm [95](Figure 3.1(a)). Given the close proximity of the signals, we assume that the regions of interest (ROIs) containing each signal location have similar background characteristics (e.g., local breast density, spatial correlations among the breast tissues). A sufficiently large ROI, covering all 4 possible signal locations, was directly extracted from a single run of CLB simulation as a synthetic multifocal case (Table 3.2). We used power law coefficient β of anatomical noise power spectrum (NPS) [34, 90] and spatial covariance $\mathbf{K}$ of image background to describe the variations of the image texture in both frequency and spatial domains. Previous studies (e.g., [34, 79, 118]) show that β increases with larger fraction of fibroglandular tissue, and $\mathbf{K}$ complements β by depicting the distribution of linear structures (e.g., ducts, vessels lead to highly-correlated points that show as lines in $\mathbf{K}$).

(ii). For multicentricity, two or more quadrants are involved, and tumors are more physically separated (Figure 3.1(a)). The ROIs representing each signal location tend to have different local background statistics [110, 120, 87]. To generate synthetic multicentric cases, we varied the parameters of CLB simulations, and extracted 4 ROI ensembles that

Table 3.2: CLB parameters: $\bar{K}$ is the mean number of clusters, $\bar{N}$ is the mean number of blobs in each cluster, L_x and L_y are the characteristic lengths in x and y direction, a, b and σ_ϕ are adjustable coefficients. The ratio of $\frac{L_x}{L_y}$, and the value of b has a small effect on the slope of NPS (i.e., β), while the value of a has a direct impact on the slope: the larger a, the smaller the slope. They are set to control the degree of fibrous appearance of the image.

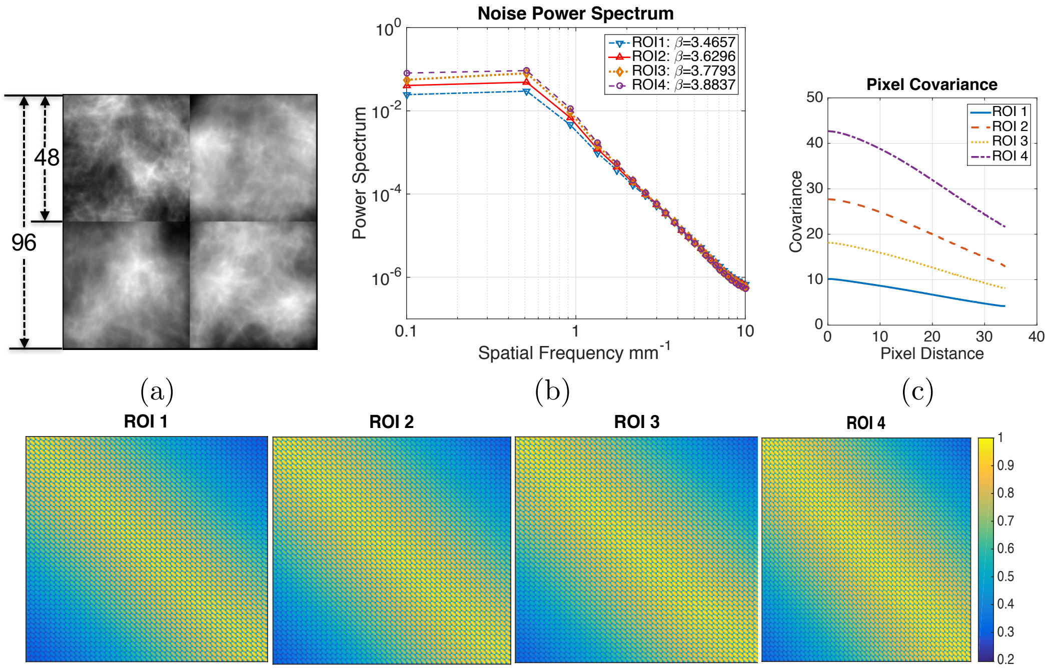
cases		$\bar{K}$	$\bar{N}$	L_x	L_y	a	b	σ_ϕ
multifocal		150	20	5	2	2.1	0.5	12
multicentric	ROI1	80	10	4	4	2.5	0.5	10.5
	ROI2	100	12	3	6	2.3	0.5	11
	ROI3	125	15	4	3	2.3	0.5	11.5
	ROI4	150	20	5	2	2.1	0.5	12

differ in β and $\mathbf{K}$ (Figure 3.2(a)-(d)). The spatial distributions of pixel intensity within each of the ROIs, and the correlations between the pixels were noticeably different. Among the ROIs, ROI 1 had the shortest correlation length, but the largest variation in intensity along the dominant direction of intensity changing, while ROI 4 had the most smoothly-distributed intensity, but the longest and strongest correlations. ROI 2 and 3 were in between, closer to ROI 1 and 4, respectively (Figure 3.3). The parameters (Table 3.2) were chosen to mimic the typical appearances of mammograms, ranging from fatty to dense breasts regions.

3.2.2.2 Synthetic breast lesions

We simulated two types of signals as the lesions: 1) circular Gaussian signals (Figure 3.1(c)), and 2) signals of realistic shapes (i.e., oval, lobulated, and irregular) (Figure 3.1(d)). For the second set, we controlled the degree of irregularity, N_{deg}, in signal shapes by adding N_{deg} randomly weighted sinusoidal waves. The waves were of frequency 1 to N_{deg} with random phases in polar coordinates. Then we blurred the signals in Cartesian coordinates with circular Gaussian masks. The contrast of the signals was set at approximately 10% relative to the mean background (i.e., the area under the receiver operating characteristic curve for an ideal observer to detect a single signal would be slightly higher than 0.80 [145]). The number of synthetic lesions (signals) in each image was sampled from a Poisson distribution, with a mean of 1. For the negligible probability (0.0037) that the number of signals was over 4, we treated these cases as 4-signals present. This truncated Poisson distribution was similar to the typical distribution of the number of tumors detected in patients with breast cancer [39]. As a result, each signal-present image has one or more signals present, and each signal is at a randomly assigned location chosen from the four pre-defined signal locations.

We simplified the problem by applying the model only to extracted ROIs, as prior research (e.g., [113, 114, 25]) has shown that carefully chosen ROIs reflect the relevant statistical properties of the images that are used for detecting signals. Each dataset included 15000 images with the resolution of 96×96 pixels: 10000 images for training the model observer, and the other 5000 as the testing set for evaluations.



(a) (b) (c)

(d)

Figure 3.2: (a) Example multicentric case with 4 ROIs generated from 4 different runs of CLB simulations: each ROI is of size 48×48 pixels and the image is of size 96×96 pixels; (b) the NPS of the four ROIs; (c) Covariance between the center pixel $(24, 24)$ and all the pixels of ROI as a function of the distance between the two pixels; (d) spatial correlations between the pixels within the four ROIs. It can be observed that the ROIs differ in local background statistics in terms of density β and spatial covariance $\mathbf{K}$. For example, ROI1 had the lowest β of 3.46, while ROI4 had the highest β of 3.88. $\mathbf{K}$ of ROI4 had relatively large (*purple* curve in (c)) and spread-out off-diagonal elements (*yellowish* regions in (d)). $\mathbf{K}$ of ROI1 was rather concentrated around the diagonal.

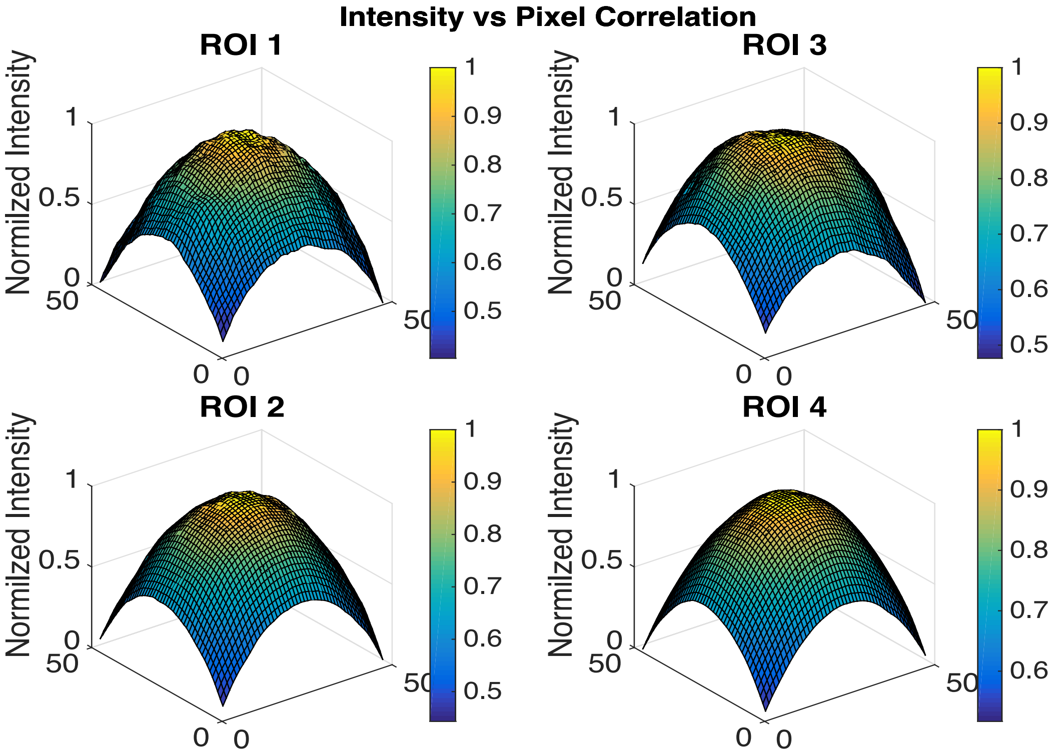


Figure 3.3: Normalized mean pixel intensity of ROIs with the correlation between the pixel $(24, 24)$ and other pixels as the colormap. It can be observed that among the ROIs, ROI 1 had the shortest correlation length, but the largest variation in intensity along the dominant direction of intensity changing, while ROI 4 had the most smoothly-distributed intensity, but the longest and strongest correlations.

3.2.3 Partial Least Squares Channels

The Hotelling observer, the optimal linear observer that maximizes the SNR, uses the mean and covariance of the image data to obtain a scalar test statistic as the decision variable [11]. Channelization is commonly used to reduce the high dimensionality of medical imaging data to make the problem solvable (e.g., [101, 114, 54, 107]). For this study, we are particularly interested in efficient channels that consider signal and background data statistics in the task. Among possible choices for channelization, PLS channels have been shown to effectively extract relevant information for signal detection tasks and approach the performance of an ideal observer (e.g., [154, 101, 107, 161]), indicating potential to accurately estimate the trend of observer performance for the multi-signal detection task being investigated. The idea of the PLS approach is to seek a small set of channels $\mathbf{T}$ that maximize the covariance between the image $\mathbf{X}$ and the decisions/truth labels $\mathbf{Y}$ [154]. As PLS channels are estimated directly from the images (without prior knowledge of the background or the signals), they are adaptive to complex image backgrounds and signal statistics. This is one of the main advantages of PLS channels over other choices of efficient channels (e.g., LG channels only work with well defined rotationally symmetric signals such as Gaussian functions can model [61]).

SIMPLS is an implementation of PLS that works well with both univariate label $\mathbf{Y}^{\mathrm{I}}$ and multivariate label $\mathbf{Y}^{\mathrm{L}}$ [42]. We adapt SIMPLS to de-

Algorithm 1 PLS channel generation

```
$\mathbf{X}_0 = \mathbf{X} - \bar{\mathbf{X}};\ \mathbf{X}_0 = \mathbf{X}_0'$
$\mathbf{Y}_0 = \mathbf{Y} - \bar{\mathbf{Y}};\ \mathbf{Y}_0 = \mathbf{Y}_0'$
$\mathbf{S} = \mathbf{X}_0'\mathbf{Y}_0$
for $i = 1, 2, ..., N_c$ do
    compute $\mathbf{r}_i$: the dominant eigenvector of SVD of $\mathbf{S}$
    $\mathbf{t}_i = \mathbf{X}_0\mathbf{r}_i$
    $\mathbf{t}_i = \frac{\mathbf{t}_i}{||\mathbf{t}_i||}$
    $\mathbf{T}_i = \frac{\mathbf{r}_i}{||\mathbf{t}_i||}$
    $\mathbf{v}_i = \mathbf{X}_0\mathbf{t}_i$
    for j=1,2,...i-1 do
        $\mathbf{v}_j = \mathbf{V}_{\{j\}}$
        $\mathbf{v}_i = \mathbf{v}_i - (\mathbf{v}_j'\mathbf{v}_i)\mathbf{v}_j$
    end for
    $\mathbf{v}_i = \frac{\mathbf{v}_i}{||\mathbf{v}_i||}$
    store $\mathbf{v}_i$ into $\mathbf{V}$ as the $i^{th}$ column: $\mathbf{V}_{\{i\}} = \mathbf{v}_i$
    $\mathbf{S} = \mathbf{S} - \mathbf{v}_i(\mathbf{v}_i'\mathbf{S})$
    $\mathbf{S} = \mathbf{S} - \mathbf{V}_{\{1:i\}}(\mathbf{V}_{\{1:i\}}'\mathbf{S})$
end for
```

velop a function, denoted as **PLS**, for generating PLS channels that are good candidates as potentially highly efficient in terms of approximating the performance of ideal observers in multiple-signal detection task [151]. The details of **PLS** are listed in Algorithm 1. With different input truth label $\mathbf{Y}$ (i.e., image-level label $\mathbf{Y}^{\mathrm{I}}$, location-specific label $\mathbf{Y}^{\mathrm{L}}$ and single-location label $\mathbf{Y}^{\mathrm{L}}(j)$, $j = 1, 2, .., N_l$), three sets of PLS channels (i.e., image-level channels $\mathbf{T}_{\mathrm{img}}$, location-level channels $\mathbf{T}_{\mathrm{loc}}$, and single-location channels $\mathbf{T}_{\mathrm{one\_loc}}$) are generated (Table 3.3). Example sets of the trained PLS channels with circular Gaussian signals are shown in Figure 3.4. It can be observed that:

(i). Signal contrasts in $\mathbf{T}_{\mathrm{img}}$ (Figure 3.4(a)) reflect signal prevalence at different locations.

Table 3.3: Notation of PLS channels

Notation	Dimension	Comments
$\mathbf{T}_{\text{img}}$	$N_p \times N_c$	$\mathbf{T}_{\text{img}} = \mathbf{PLS}(\mathbf{X}, \mathbf{Y}^{\text{I}})$
$\mathbf{T}_{\text{loc}}$	$N_p \times N_c$	$\mathbf{T}_{\text{loc}} = \mathbf{PLS}(\mathbf{X}, \mathbf{Y}^{\text{L}})$
$\mathbf{T}_{\text{one\_loc}}$	$N_p \times N_c \times N_l$	For $j = 1, 2, ..., N_l$, $\mathbf{T}_{\text{one\_loc}}\{j\} = \mathbf{PLS}(\mathbf{X}, \mathbf{Y}^{\text{L}}(j))$

(ii). $\mathbf{T}_{\text{loc}}$ (Figure 3.4(b)) captures both signal prevalence and interactions between signal locations, which can be used as the common set of channels to generate both image-level and location-specific decisions.

(iii). $\mathbf{T}_{\text{one\_loc}}$ (Figure 3.4(c)) is specific to one particular signal location, and it does not incorporate the spatial correlations between the signals and backgrounds.

Note that $\mathbf{T}_{\text{loc}}$ is of particular interest as it can be used to generate both image-level and location-specific decisions, and it is universal to all possible signal locations. This also leads to a key advantage of $\mathbf{T}_{\text{loc}}$ over $\mathbf{T}_{\text{one\_loc}}$, because the model observer with $\mathbf{T}_{\text{loc}}$ only uses $\frac{1}{N_l}$ times the total number of channels used with $\mathbf{T}_{\text{one\_loc}}$.

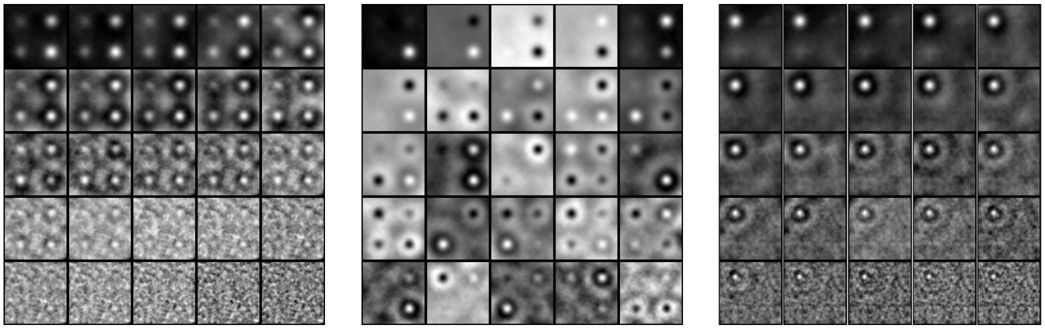
3.2.4 Laguerre Gauss Channels

The model observer with the modified LG channels served as the benchmark. LG channels [54], the product of Laguerre polynomials and Gaussian functions, are often used in task-based assessments of medical image quality (e.g., [158, 32]). One advantage of LG channels is that LG channels are based on an analytical expression, so the quality of LG channels is independent of the number of available training samples [61]. Inspired by the counterpart PLS channels, we constructed three sets of modified LG channels. The width parameter of the symmetric LG channels was chosen to fit the spread of signals and correlation length of CLB [160, 54]..

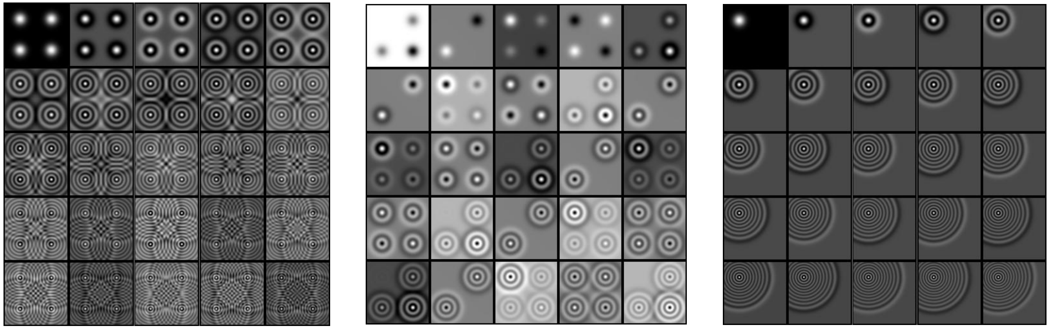
(i). $\mathbf{T}_{\text{img}}$: Four sets of LG channels centered at four signal locations were added together then normalized as $\mathbf{T}_{\text{img}}$ (Figure 3.4(d)).

(ii). $\mathbf{T}_{\text{loc}}$: Given $N_l = 4$ and at most one signal presents at each of the locations, there were a total of 15 possible scenarios of signal-presence in the image (i.e., 4 scenarios with one signal, 6 scenarios with two signals, 4 scenarios with three signals, 1 scenario with all four signals). We first sorted the empirical probability p that each possible scenario occurred in the training set. For each scenario with $p > 0.025$, we added LG channels centered at each of the signal-present locations. Starting with 0^{th}-order LG channels, we looped over all scenarios, in the descending order of p, before repeating the steps with 1^{st}-order LG channels. An orthonormal set of channels was extracted as the final $\mathbf{T}_{\text{loc}}$ by orthonormalization (Figure 3.4(e)).

(iii). $\mathbf{T}_{\text{one\_loc}}$: A set of LG channels centered at the $j^{\text{th}}(j = 1, 2, , N_l)$ signal location was denoted as $\mathbf{T}_{\text{one\_loc}}\{j\}$ (Figure 3.4(f)).



(a) PLS $\mathbf{T}_{\text{img}}$ (b) PLS $\mathbf{T}_{\text{loc}}$ (c) PLS $\mathbf{T}_{\text{one\_loc}}\{1\}$



(d) LG $\mathbf{T}_{\text{img}}$ (e) LG $\mathbf{T}_{\text{loc}}$ (f) LG $\mathbf{T}_{\text{one\_loc}}\{1\}$

Figure 3.4: (a)-(c): First 25 PLS channels of $\mathbf{T}_{\text{img}}$, $\mathbf{T}_{\text{loc}}$, $\mathbf{T}_{\text{one\_loc}}\{1\}$. (d)-(f): First 25 LG channels $\mathbf{T}_{\text{img}}$, $\mathbf{T}_{\text{loc}}$, $\mathbf{T}_{\text{one\_loc}}\{1\}$. For each subfigure, 1st row shows the 1st (*left*) to 5th (*right*) channel, 2nd row shows the 6th (*left*) to 10th (*right*) channel and etc.

3.2.5 $\mathbf{w}_{\mathbf{CHO}}$ Template

A linear discriminant is commonly used to separate two classes of objects [61]. Generally, the w_{CHO} template $\mathbf{w}$ for computing a linear discriminant is obtained by:

$$\mathbf{w} = \mathbf{K}_{\mathbf{V}}^{-1}\bar{\mathbf{V}} \tag{3.1}$$

where $\mathbf{V} = \mathbf{T}^T\mathbf{X}$ (^{T} indicates transpose) is channelized image; $\bar{\mathbf{V}} = \bar{\mathbf{V}}_1 - \bar{\mathbf{V}}_0$ is mean channelized signal(s); and $\mathbf{K}_{\mathbf{V}} = \frac{1}{2}(\mathbf{K}_{\mathbf{V}_1} + \mathbf{K}_{\mathbf{V}_0})$ is mean covariance

Table 3.4: Notation of $\mathrm{w}_{\mathrm{CHO}}$ templates

Notation	**Dimension**	**Comments**
$\mathbf{w}^{\mathrm{I}}_{\mathrm{img}}$	$N_c \times 1$	$\mathbf{w}^{\mathrm{I}}_{\mathrm{img}} = \mathbf{CHOtemp}(\mathbf{T}_{\mathrm{img}}, \mathbf{X}, \mathbf{Y}^{\mathrm{I}})$
$\mathbf{w}^{\mathrm{I}}_{\mathrm{loc}}$	$N_c \times 1$	$\mathbf{w}^{\mathrm{I}}_{\mathrm{loc}} = \mathbf{CHOtemp}(\mathbf{T}_{\mathrm{loc}}, \mathbf{X}, \mathbf{Y}^{\mathrm{I}})$
$\mathbf{w}^{\mathrm{L}}_{\mathrm{loc}}$	$N_c \times N_l$	$\mathbf{w}^{\mathrm{L}}_{\mathrm{loc}} = \mathbf{CHOtemp}(\mathbf{T}_{\mathrm{loc}}, \mathbf{X}, \mathbf{Y}^{\mathrm{L}})$
$\mathbf{w}^{\mathrm{L}}_{\mathrm{one\_loc}}$	$N_c \times N_l$	For $j = 1,..,N_l, \mathbf{w}^{\mathrm{L}}_{\mathrm{one\_loc}}(j) = \mathbf{CHOtemp}(\mathbf{T}_{\mathrm{one\_loc}}\{j\}, \mathbf{X}, \mathbf{Y}^{\mathrm{L}}(j))$

matrix of $\mathbf{V}_1$ and $\mathbf{V}_0$. With three sets of $\mathbf{T}$, $\mathbf{V}$ are classified into $\mathbf{V}_1$ and $\mathbf{V}_0$ differently based on decisions of interest (i.e., $\mathbf{Y}^{\mathrm{I}}$ or $\mathbf{Y}^{\mathrm{L}}$):

(i). For $\mathbf{w}^{\mathrm{I}}_{\mathrm{img}}$ and $\mathbf{w}^{\mathrm{I}}_{\mathrm{loc}}$, $\mathbf{V}_1$ are the columns of $\mathbf{V}$ that the corresponding element of $\mathbf{Y}^{\mathrm{I}}$ is 1 (i.e., the images with at least one signal), and $\mathbf{V}_0$ are the columns that $\mathbf{Y}^{\mathrm{I}}$ is 0.

(ii). For $\mathbf{w}^{\mathrm{L}}_{\mathrm{loc}}(j)$ and $\mathbf{w}^{\mathrm{L}}_{\mathrm{one\_loc}}(j)$, $\mathbf{V}_1(j)$ are the columns that $\mathbf{Y}^{\mathrm{L}}(j)$ is 1, and $\mathbf{V}_0(j)$ are the columns that $\mathbf{Y}^{\mathrm{L}}(j)$ is 0, where $j = 1, 2, .., N_l$

Hence, **CHOtemp** as a function of $\mathbf{T}$, $\mathbf{X}$ and $\mathbf{Y}$ trains four different sets of $\mathbf{w}$, shown in Table 3.4.

3.2.6 Decision Variables

Decision variables S for making detection decisions are computed: $S = \mathbf{w}^T(\mathbf{T}^T\mathbf{X})$. The higher the score is, the more likely the signal is present. Specifically, with four different combinations of $\mathbf{T}$ and $\mathbf{w}$, we have four sets of decision variables for each image:

(i). $S^{\mathrm{I}}_{\mathrm{img}} = (\mathbf{w}^{\mathrm{I}}_{\mathrm{img}})^T(\mathbf{T}^T_{\mathrm{img}}\mathbf{X})$

(ii). $S^{\mathrm{I}}_{\mathrm{loc}} = (\mathbf{w}^{\mathrm{I}}_{\mathrm{loc}})^T(\mathbf{T}^T_{\mathrm{loc}}\mathbf{X})$

(iii). $\mathbf{S}^{\mathrm{L}}_{\mathrm{loc}} = (\mathbf{w}^{\mathrm{L}}_{\mathrm{loc}})^T(\mathbf{T}^T_{\mathrm{loc}}\mathbf{X})$

(iv). $\mathbf{S}^{\mathrm{L}}_{\mathrm{one\_loc}} = (\mathbf{w}^{\mathrm{L}}_{\mathrm{one\_loc}})^T(\mathbf{T}^T_{\mathrm{one\_loc}}\mathbf{X})$

For each image, $S^{\mathrm{I}}_{\mathrm{img}}$ and $S^{\mathrm{I}}_{\mathrm{loc}}$ are scalars, representing image-level scores for making image-level decisions. $\mathbf{S}^{\mathrm{L}}_{\mathrm{loc}}$ and $\mathbf{S}^{\mathrm{L}}_{\mathrm{one\_loc}}$ are vectors of N_l scores, representing location-specific scores for making both image-level and location-specific decisions. One extra step was used to convert $\mathbf{S}^{\mathrm{L}}_{\mathrm{loc}}$ and $\mathbf{S}^{\mathrm{L}}_{\mathrm{one\_loc}}$ into scalar image-level scores, denoted as $\mathbf{S}^{\mathrm{I}}_{\mathrm{loc}}$ and $\mathbf{S}^{\mathrm{I}}_{\mathrm{one\_loc}}$, respectively. The idea was to apply a second-stage linear decision template [114]: $S^{\mathrm{I}} = (\mathbf{w}^{\mathrm{I}})^T\mathbf{S}^{\mathrm{L}}$, where $\mathbf{w}^I = \mathbf{K}_{\mathbf{S}}^{-1}\bar{\mathbf{S}}$; $\bar{\mathbf{S}} = \bar{\mathbf{S}}_1 - \bar{\mathbf{S}}_0$ is mean difference in $\mathbf{S}$ under signal-present and signal-absent hypotheses; $\mathbf{K}_{\mathbf{S}} = \frac{1}{2}(\mathbf{K}_{\mathbf{S}_1} + \mathbf{S}_{\mathbf{V}_0})$ is mean covariance of $\mathbf{S}$. $\mathbf{S}^{\mathrm{L}}$ can be either $\mathbf{S}^{\mathrm{L}}_{\mathrm{loc}}$ or $\mathbf{S}^{\mathrm{L}}_{\mathrm{one\_loc}}$.

3.2.7 Evaluations

3.2.7.1 Figure of merit

For image-level decisions, SNR was computed to quantify the ability of the model observer in making binary decisions [61]. SNR is defined as: $\mathrm{SNR} = \frac{<S>_{\mathbf{H}_1} - <S>_{\mathbf{H}_0}}{\sqrt{\frac{1}{2}(\sigma^2_{\mathbf{H}_0} + \sigma^2_{\mathbf{H}_1})}}$, where $< S >_{\mathbf{H}_i}$ $(i = 0, 1)$ and $\sigma^2_{\mathbf{H}_i}(i = 0, 1)$ is the mean and variance of S under signal-present ($\mathbf{H}_1$) or signal-absent hypothesis ($\mathbf{H}_0$), respectively. Corresponding receiver operating characteristics (ROC) curves were also plotted to provide additional information for assessing image-level detection performance.

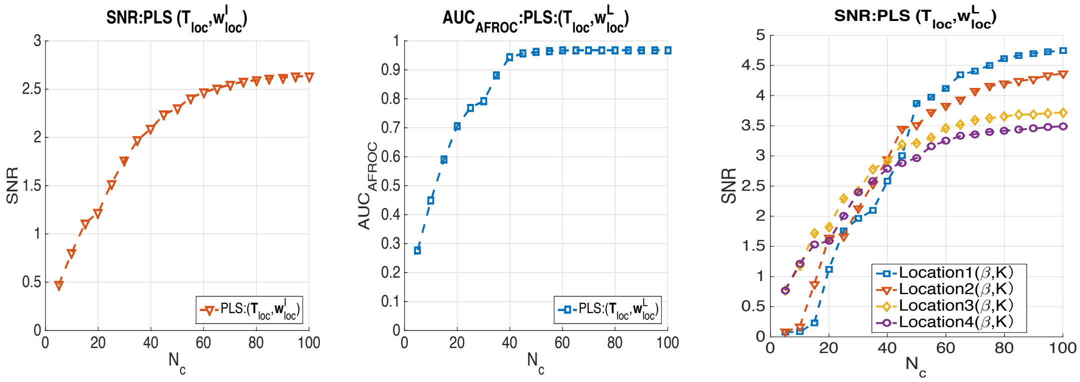
For location-specific decisions, location-specific SNRs were computed to assess the accuracy of the model observer at each of the signal locations. Moreover, an alternative free-response receiver operating characteristics curve (AFROC) [30] was plotted. AFROC is befitting here as we assume that there is no additional penalty for more than one false positive mark in a signal-absent case [157]. A corrected localized signal was counted if the decision score at a signal location was above the detection threshold, and a signal was indeed present at that particular location. One false positive image was counted if any of the four locations was a false positive, where a false positive was defined as the score at a signal location was above the threshold, but no signal was present at that location. The area under the curve ($\mathrm{AUC}_{\mathrm{AFROC}}$) in the range of $[0, 1]$ was estimated as the FOM [30].

3.2.7.2 Benchmark single-signal CHOs

Two existing CHO models originally built for single-signal detections were extended as the benchmark for multi-signal detection tasks. The two models both used a two-stage design [114], but differed in the first stage for computing ROI-specific scores: Model 1 only trained one CHO by mixing all the four ROIs together, while Model 2 treated each ROI ensemble as one dataset, and trained four CHOs separately (i.e., one set of PLS channels and CHO template for each of the four ROIs). Hence, none of the sets of PLS channels captured the interactions among the ROIs. The second stage of both Model 1 and 2 trained a CHO to convert ROI-specific scores into image-level scores. We denoted the multi-signal model observers using PLS $\mathbf{T}_{\text{loc}}$ and $\mathbf{T}_{\text{one\_loc}}$ channels as Model 3 and 4, respectively. Unlike Model 1 and 2, they assess the entire image altogether and incorporate the interactions among local statistics of the ROIs.

3.3 Results

In the pilot study [151], we have shown that for multifocal cases, the model observers with a small number of channels could reliably detect multiple circular Gaussian signals and their locations. In another pilot study [150], we have shown that local anatomic variations may significantly affect the detection of multiple breast lesions on mammogram. In the extended study [152], we further assess the efficacy of the model observers with the realism of signals and complexity of background statistics. This section highlights the evaluation results reported in [152], and is organized as follows: Section 3.3.1 reports the FOM results with synthetic multicentric cases and irregularly shaped signals. Section 3.3.2 shows the benefits of incorporating multi-location interactions. Section 3.3.3 studies the impact of sample size on the model observer performance. Counterpart results with circular Gaussian signals in multicentric cases are summarized in Appendix A.2.

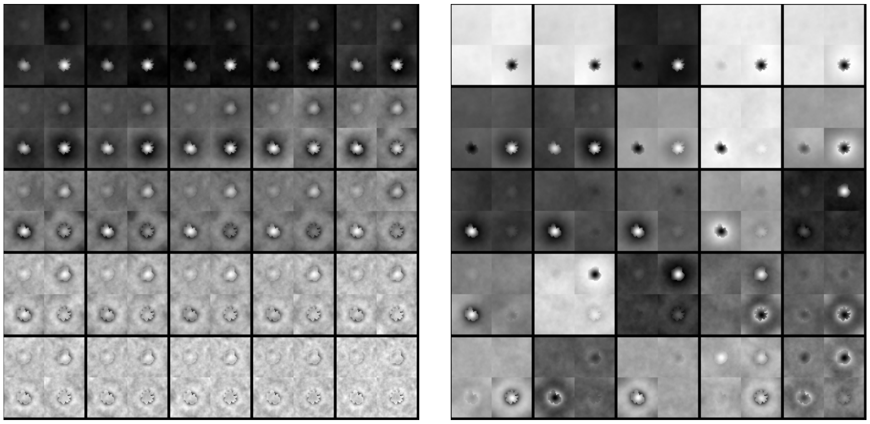


(a) Image-level SNR (b) $\text{AUC}_{\text{AFROC}}$ (c) Location-specific SNRs

Figure 3.5: (a) Image-level SNR: PLS ($\mathbf{T}_{\text{loc}}, \mathbf{w}^{\text{I}}_{\text{loc}}$); (b) $\text{AUC}_{\text{AFROC}}$: PLS ($\mathbf{T}_{\text{loc}}, \mathbf{w}^{\text{L}}_{\text{loc}}$); (c) Location-specific SNRs: PLS ($\mathbf{T}_{\text{loc}}, \mathbf{w}^{\text{L}}_{\text{loc}}$). These results show that the model observers were able to perform the multi-signal detection tasks with accuracy.

3.3.1 Results with Signals of Realistic Shapes

The FOM results validate that the model observers are able to perform the multi-signal detection tasks with accuracy. For illustration, Figure 3.5 shows the FOMs with PLS $\mathbf{T}_{\text{loc}}$ channels as a function of N_c. With a larger N_c, the image-level SNR (Figure 3.5(a)), $\text{AUC}_{\text{AFROC}}$ (Figure 3.5(b)), and the location-specific SNRs (Figure 3.5(c)) all increased, and finally reached performance plateaus when N_c is sufficiently large (e.g., $\text{AUC}_{\text{AFROC}}$ is close to 1 when $N_c = 40$ in Figure 3.5(b)). Similar but not exactly same performance trends are observed with the other designs of the model observer (i.e., PLS $\mathbf{T}_{\text{img}}$ channels, PLS $\mathbf{T}_{\text{one\_loc}}$ channels, LG channels) (Appendix A.1 Figure A.1). The similarities and differences will be discussed in detail next.

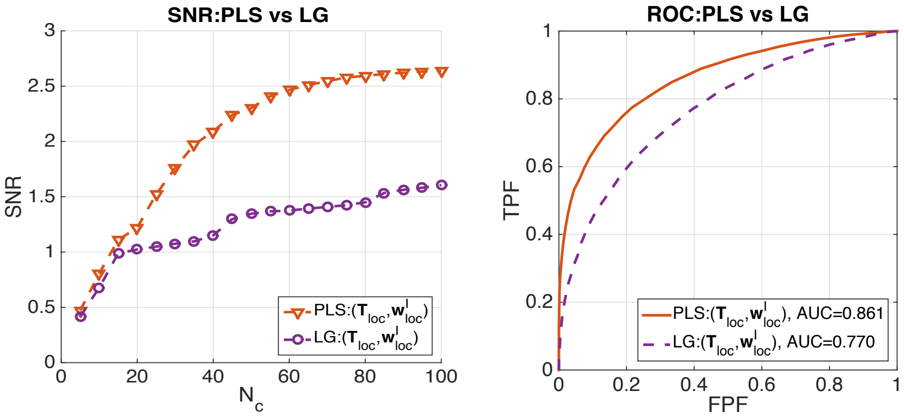


(a) PLS $\mathbf{T}_{\mathrm{img}}$ (b) PLS $\mathbf{T}_{\mathrm{loc}}$

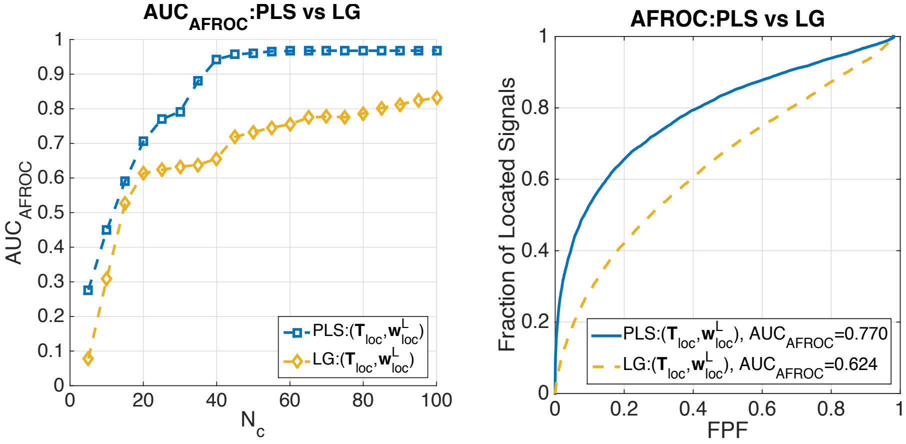
Figure 3.6: (a)-(b) shows the first 25 PLS channels of $\mathbf{T}_{\mathrm{img}}$, $\mathbf{T}_{\mathrm{loc}}$ in multicentric cases with realistic signals.The channels captured signal characteristics at different locations.

3.3.1.1 PLS channels versus LG channels

The three sets of PLS channels (e.g., $\mathbf{T}_{\mathrm{img}}$ in Figure 3.6(a), $\mathbf{T}_{\mathrm{loc}}$ in Figure 3.6(b)) captured signal characteristics (e.g., shapes) at different locations, while the counterpart LG channels (Figure 3.4(d)-(e)) failed to adapt to the signals. Thus, the model observers with PLS channels significantly outperformed those with LG channels (e.g., image-level SNR with $\mathbf{T}_{\mathrm{loc}}$ channels in Figure 3.7(a), $\mathrm{AUC}_{\mathrm{AFROC}}$ with $\mathbf{T}_{\mathrm{loc}}$ channels in Figure 3.7(c), and the corresponding example ROC and AFROC curves achieved with $N_c = 25$ channels in Figure 3.7(b)(d)). This difference reflects the limitation of LG channels that they only work for stationary background and rotationally symmetric signals [61]. The LG channels failed to capture all essential information for signal detection when realistic signals and background were involved in the detection tasks.



(a) Image-level SNR (b) ROC curve, $N_c = 25$



(c) $\mathrm{AUC}_{\mathrm{AFROC}}$ (d) AFROC curve, $N_c = 25$

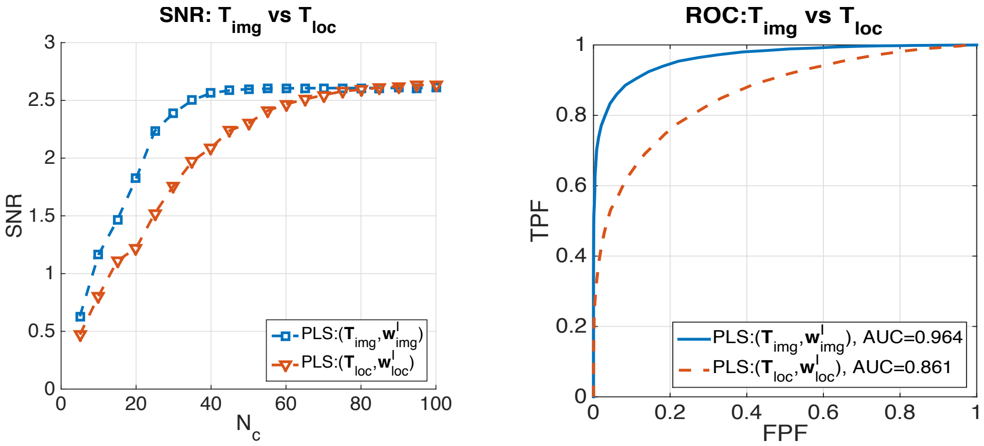
Figure 3.7: (a) Image-level SNR: PLS $(\mathbf{T}_{\mathrm{loc}}, \mathbf{w}^{\mathrm{I}}_{\mathrm{loc}})$ (*red*) versus LG $(\mathbf{T}_{\mathrm{loc}}, \mathbf{w}^{\mathrm{I}}_{\mathrm{loc}})$ (*purple*); (b) Example image-level ROC curves achieved with 25 channels (i.e., $N_c = 25$): PLS $(\mathbf{T}_{\mathrm{loc}}, \mathbf{w}^{\mathrm{I}}_{\mathrm{loc}})$ (*red*) versus LG $(\mathbf{T}_{\mathrm{loc}}, \mathbf{w}^{\mathrm{I}}_{\mathrm{loc}})$ (*purple*); (c) $\mathrm{AUC}_{\mathrm{AFROC}}$: PLS $(\mathbf{T}_{\mathrm{loc}}, \mathbf{w}^{\mathrm{L}}_{\mathrm{loc}})$ (*blue*) and LG $(\mathbf{T}_{\mathrm{loc}}, \mathbf{w}^{\mathrm{L}}_{\mathrm{loc}})$ (*yellow*); (d) Example AFROC curves achieved with 25 channels ($N_c = 25$): PLS $(\mathbf{T}_{\mathrm{loc}}, \mathbf{w}^{\mathrm{L}}_{\mathrm{loc}})$ (*blue*) and LG $(\mathbf{T}_{\mathrm{loc}}, \mathbf{w}^{\mathrm{L}}_{\mathrm{loc}})$ (*yellow*); It can be observed that PLS channels significantly outperformed the counterpart LG channels.

3.3.1.2 Image-level decisions: SNR

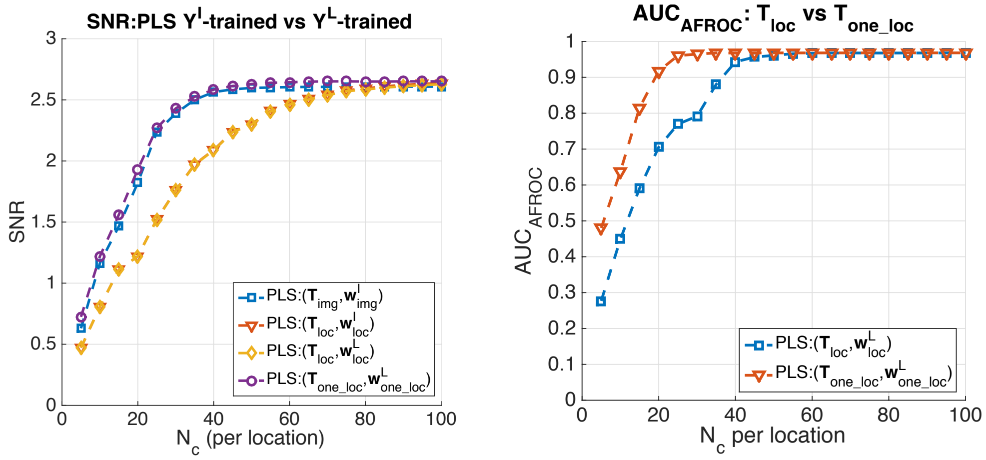
When the number of channels N_c is small, $\mathbf{T}_{\text{img}}$ PLS channels are superior to $\mathbf{T}_{\text{loc}}$ PLS channels (Figure 3.8(a)). However, when the N_c is ≥ 60, the SNR of $\mathbf{T}_{\text{loc}}$ is approximately the same as the SNR of $\mathbf{T}_{\text{img}}$. Example ROC curves achieved at $N_c = 25$ (Figure 3.8(b)) also shows the advantage of $\mathbf{T}_{\text{img}}$ over $\mathbf{T}_{\text{loc}}$ when N_c is small. Whereas $\mathbf{T}_{\text{img}}$ only works for image-level decisions, $\mathbf{T}_{\text{loc}}$ is versatile, and can be used to generate both image-level and location-specific decisions. The SNRs of $\mathbf{Y}^{\text{L}}$-trained $\mathbf{T}$ and $\mathbf{w}$ (*blue* and *red* curves in Figure 3.8(c)) were similar to the SNRs of $\mathbf{Y}^{\text{I}}$-trained $\mathbf{T}$ and $\mathbf{w}$ (*purple* and *yellow* curves in Figure 3.8(d)), but a larger total N_c is required (e.g., a total of $N_c \times 4$ channels used in $\mathbf{T}_{\text{one\_loc}}$).

3.3.1.3 Location-level decisions: $\text{AUC}_{\text{AFROC}}$

The performance, measured using $\text{AUC}_{\text{AFROC}}$, of $\mathbf{T}_{\text{loc}}$ PLS channels is comparable to that of $\mathbf{T}_{\text{one\_loc}}$ PLS channels when N_c per location ≥ 40 (Figure 3.8(c)). This shows a huge reduction in the total N_c (e.g., using 40 $\mathbf{T}_{\text{loc}}$ channels reached equivalent performance as using 160 $\mathbf{T}_{\text{one\_loc}}$ channels). This is especially valuable when N_l is large (e.g., 20) in other multi-signal detection tasks. Hence, when the total N_c to be used is limited (e.g., computational power, memory), location-universal $\mathbf{T}_{\text{loc}}$ may perform considerably better than location-specific $\mathbf{T}_{\text{one\_loc}}$.



(a) SNR: $\mathbf{T}_{\text{img}}$ vs $\mathbf{T}_{\text{loc}}$ (b) ROC curve, $N_c = 25$: $\mathbf{T}_{\text{img}}$ vs $\mathbf{T}_{\text{loc}}$



(c) SNR: $\mathbf{Y}^{\text{I}}$-trained vs $\mathbf{Y}^{\text{L}}$-trained (d) $\text{AUC}_{\text{AFROC}}$: $\mathbf{T}_{\text{loc}}$ vs $\mathbf{T}_{\text{one\_loc}}$

Figure 3.8: (a) Image-level SNR: PLS $(\mathbf{T}_{\text{img}}, \mathbf{w}^{\text{I}}_{\text{img}})$ (*blue*) versus PLS $(\mathbf{T}_{\text{loc}}, \mathbf{w}^{\text{I}}_{\text{loc}})$ (*red*). When N_c is small, $\mathbf{T}_{\text{img}}$ are superior to $\mathbf{T}_{\text{loc}}$, but when N_c is ≥ 60, the SNR of $\mathbf{T}_{\text{loc}}$ is approximately the same as the SNR of $\mathbf{T}_{\text{img}}$. (b) Example ROC curve achieved with 25 channels. As $N_c = 25$ is relatively small, $\mathbf{T}_{\text{loc}}$ (*red*) slightly underperforms $\mathbf{T}_{\text{img}}$ (*blue*). (c) Image-level SNR: $\mathbf{Y}^{\text{I}}$-trained PLS (*blue, red*) versus $\mathbf{Y}^{\text{L}}$-trained PLS (*yellow, purple*). The SNRs of $\mathbf{Y}^{\text{L}}$-trained were similar to the SNRs of $\mathbf{Y}^{\text{I}}$-trained. (c) $\text{AUC}_{\text{AFROC}}$: PLS $(\mathbf{T}_{\text{loc}}, \mathbf{w}^{\text{L}}_{\text{loc}})$ (*blue*) versus PLS $(\mathbf{T}_{\text{one\_loc}}, \mathbf{w}^{\text{L}}_{\text{one\_loc}})$ (*red*). The $\text{AUC}_{\text{AFROC}}$ of $\mathbf{T}_{\text{loc}}$ is comparable to that of $\mathbf{T}_{\text{one\_loc}}$ when N_c per location ≥ 40.

3.3.1.4 Location-level decisions: location-specific SNRs

The location-specific SNRs of the model observer were significantly different across the four signal locations (Figure 3.5(c)), when N_c is sufficiently large (e.g., 50 per location of $\mathbf{T}_{\text{one\_loc}}$ PLS channels). This agrees with our expectation that the four signal locations in multicentric cases with different local background statistics (i.e., β and $\mathbf{K}$) would lead to different levels of detectability. Similarly, the model with $\mathbf{T}_{\text{loc}}$ used a much fewer channels than $\mathbf{T}_{\text{one\_loc}}$, thus will require less training for reaching the same level of location-specific SNRs.

(i). *Impact of* β: For each of the four ROIs, we estimated the power law exponent β from the NPS curves (Figure 3.2(b)). ROI1, with the lowest β of 3.46, had the highest SNR, while ROI4, with the highest β of 3.88, has the lowest SNR (Figure 3.5(c)). This agrees with the general performance trend found in prior ideal model studies (e.g., [27, 34, 118]) and human observer studies on 2D mammograms (e.g., [26, 85]): a higher β of is associated with increasing breast density and decreasing lesion detectability. This shows the potential of the model observers to adapt to the variations in anatomical backgrounds.

(ii). *Impact of* $\mathbf{K}$: The spatial covariance $\mathbf{K}$ of ROI4 had relatively large and spread-out off-diagonal elements (Figure 3.2(d)), which implied long correlation length and strong correlations among the pixels. $\mathbf{K}$ of ROI1 was rather concentrated around the diagonal. Hence, ROI4 and ROI1 had the lowest and highest SNR, respectively (Figure 3.5(c)).

3.3.2 Interactions among Signal Locations

The model observers assess the entire image altogether, and incorporate the interactions among ROIs. This is beneficial in multi-signal detection tasks, as signal presence at one location may alter the detection at other locations (e.g., concurrence), and local statistics of the ROIs (e.g., different β's and $\mathbf{K}$'s as the result of CLB simulations) impacts the trends of observer performance (Section 3.3.1). To demonstrate the benefits of such design, we compared the performance of the four model observers (Section 3.2.7.2). Circular signals of different sizes and contrasts were used.

(i). Model 3 and 4 were consistently good when compared to Model 1 and 2 (Figure 3.9(a)-(b)).

(ii). Model 1 (*blue*), measured by image-level SNR, significantly underperformed the other three models (Figure 3.9(a)). As 1$^{\text{st}}$ stage of Model 1 only trained one CHO by mixing all the ROIs together, the ROI-specific scores were computed in a global manner, and failed to capture different local background statistics of the ROIs.

(iii). Model 2 (*red*), measured by $\text{AUC}_{\text{AFROC}}$, significantly underperformed the other three models (Figure 3.9(b)). As Model 2 trained four separate CHOs by treating each ROI ensemble as one dataset, the ROI-specific scores were computed locally (e.g., of different scales), and failed to capture the global interactions among the ROIs.

(iv). Model 2 achieved higher location-specific SNRs than Model 1, but approximately equal SNRs as Model 3 and 4. This is because Model 2 trained one CHO for each of the ROIs, but Model 1 mixed all the ROIs. The ROI scores by Model 2 were more specific for each ROI than those from Model 1.

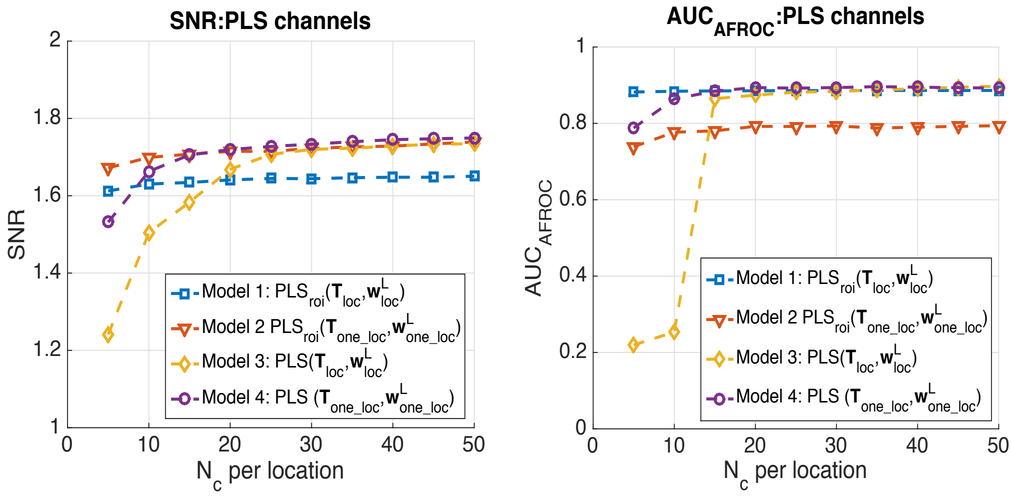
Similar results and performance trends were observed with LG channel (Appendix A.3 Figure A.3).

3.3.3 Sample Size Considerations

We assessed the model observer performance with various sizes of training sets. The goal was to understand how limited samples may be optimally used in imaging research. For illustration, we focused on the model observer using $\mathbf{T}_{\mathrm{loc}}$ PLS channels, but similar trends were seen with other designs. Irregularly shaped signals were used.

3.3.3.1 The number of PLS training images

We fixed the number of $\mathrm{w}_{\mathrm{CHO}}$ training and testing images at 5000, and varied the number of PLS training images to one of $\{100, 200, 500, 1000, 2500, 5000, 10000\}$. Each setting was repeated for five times, and the templates $\mathbf{w}$ were sufficiently trained for each setting. Despite the notable differences in the absolute performance, the trend of image-level SNR (Figure 3.10(a))



(a) Image-level SNR (b)$\mathrm{AUC_{AFROC}}$

Figure 3.9: These results of multicentric cases with circular Gaussian signals are shown to demonstrate the benefits of incorporating the interactions among signal locations in multi-signal detection tasks. (a) image-level SNR curves show that Model 1 (*blue*) significantly underperforms the other three. (b) $\mathrm{AUC_{AFROC}}$ curves show that Model 2 (*red*) significantly underperforms the other three. This is because Model 1, only training one CHO by mixing all the four ROIs together, computes the ROI-specific scores globally, while Model 2, training four separate CHOs, computes the ROI-specific scores locally.

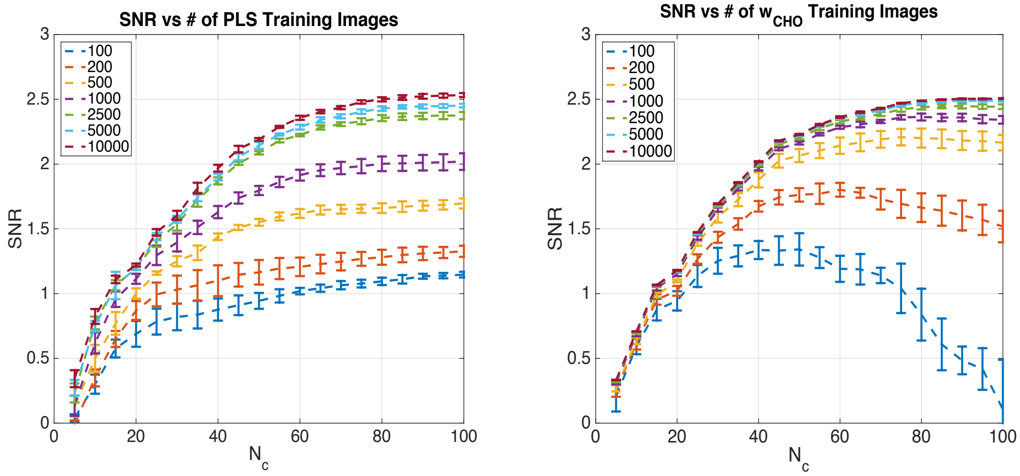
and $\mathrm{AUC_{AFROC}}$ (Appendix A.4 Figure A.4(a)) as a function of N_c was similar across different numbers of PLS training images. Hence, PLS channels, estimated from a limited number of training images, may still efficiently capture useful discriminant information for predicting the trend of accuracy in making detection decisions, even if the quality of the channels may not as good as they could possibly be.

3.3.3.2 The number of $\mathbf{w}_{\text{CHO}}$ training images

Similarly, we fixed the number of PLS training and testing images at 5000, and varied the number of w_{CHO} training images. For small numbers (e.g., 100, 200), the image-level SNR significantly degraded at a large N_c (Figure 3.10(b)), and $\text{AUC}_{\text{AFROC}}$ was substantially lower with a larger variation (Appendix A.4 Figure A.4(b)). This may be caused by inaccurate, and/or instable estimation of the channelized covariance, when such a small number of training images were used. Given that the quality of PLS channels was consistently high, this suggests that the sample size of w_{CHO} training may have a larger impact on the performance trend. Thus, if the total sample size is limited, a larger fraction of the images need be used to train the template $\mathbf{w}$.

3.4 Discussion

We have constructed novel implementation of PLS channels and LG channels for a multi-signal detection task. PLS channels are estimated directly from the images, and thus they naturally adapt to variations in signals and backgrounds. This key advantage of PLS channels is evident as shown in Section 3.3.1. However, unlike analytically derived LG channels, the quality of PLS channels highly depends on the number of channel training images. On the other hand, the results with circular Gaussian signals (Appendix A.2 Figure A.2) exhibit the effectiveness of LG channels. Fewer LG channels were needed than PLS to reach the comparable perforance plateau. However, we also need to admit the limitation of LG channels that they are only good



(a) SNR vs no. of PLS training images (b) SNR vs no. of $\mathrm{w}_{\mathrm{CHO}}$ training images

Figure 3.10: Impacts of the number of training samples on the performance of the model observers with $\mathbf{T}_{\mathrm{loc}}$ PLS channels. (a)-(b) Plots of image-level SNR as the number of PLS training images, and $\mathrm{w}_{\mathrm{CHO}}$ training images changes, respectively. The performance trend as a function of N_c was similar across different numbers of PLS training images. With smaller number of $\mathrm{w}_{\mathrm{CHO}}$ training images, image-level SNR degraded significantly at large N_c.

for such rotationally symmetric signals. Consequently, there is a substantial decrease in detection performance when realistic signals were involved in the tasks (Figure 3.7). Thus, these tradeoffs need to be carefully considered when choosing channels in practice. Moreover, other channelization methods (e.g., DoG channels, Gabor channels [61]) may be extended in a similar way as was done for LG channels (Section 3.2.4).

Anatomical structures vary across images regions. Previous studies (e.g., [154], [81]) have shown that image background is essential for efficient channel computations, and signal detections. The trends of FOMs (e.g., image-level SNR and $\mathrm{AUC}_{\mathrm{AFROC}}$) for multifocal cases [151]) and multicentric cases were generally similar. The location-specific SNRs were the same for multifocal cases, but noticeably different for multicentric cases. This suggests that local background statistics (e.g., β and $\mathbf{K}$ of a ROI) to an extent determine the detectability within the ROI. In addition, the prevalence of signals and the interactions among ROIs are also critical for making detection decisions (Section 3.3.2). Hence, $\mathbf{T}_{\mathrm{loc}}$ is highlighted as it captures all the aforementioned information. However, being universal to all possible signal locations may degrade its ability to match $\mathbf{T}_{\mathrm{img}}$ for image-level performance at small N_c, and $\mathbf{T}_{\mathrm{one\_loc}}$ for location-level performance at small N_c per location. This is more evident when the sample size available for training is limited. Section 3.3.3 reveals that the allocations of channel training and $\mathrm{w}_{\mathrm{CHO}}$ training may greatly impact the detection performance. Thus, the appropriate design of a model observer depends on both the task of clinical interest and the available resources.

When compared to prior studies on multi-signal detection, our model observers do not need assumptions of stationary or known background, or signal independence: they are adaptive to signals, and local background statistics. However, one of the limitations of the study is that we assume there are only a finite number of possible signal locations (e.g., $N_l = 4$), and these locations are fixed. To overcome this pitfall, we could add the step, similar to the two-stage design in [7], to search globally for suspicious candidate locations. Alternatively, we may simply use smaller ROIs, and increase N_l. However, how such changes may impact the model training, and model performance requires further investigations.

This study focuses on evaluating the model observer in the presence of anatomic noise. This was motivated by the fact that the variability of the anatomical background has been shown as the dominant distracting factor for many detection tasks in breast imaging (e.g., [27, 17, 29, 67]). However, prior studies (e.g., [124, 122, 119]) also suggest that quantum noise affects lesion detectability, especially for small-scale objects such as microcalcifiations. As there is increasing concern on radiation dose reduction when designing imaging system, it would be valuable to employ the model observer to investigate how increased quantum noise obtained at reduced dose may influence multi-signal detection. The study could be potentially extended by two approaches: 1). Quantum noise at varying dose levels could be simulated and added to anatomic image background. For example, the Anscombe transformation [19] converts object-dependent quantum noise in the spatial domain into object-independent Gaussian noise in the Anscombe domain. Changing the variance of Gaussian noise would lead to different quantum noise (i.e., dose levels). Without modifying the model observer design, the evaluation procedures could be repeated. 2) Covariance decomposition [158] that represents the total variance as the sums of the variance due to quantum noise, noise in the gain processes,electronic noise effects, and anatomical variations could be used. When dose level is changed, the components of total covariance affected by the changed dose could be systematically computed. In this way, not all model observer calculations have to be repeated. Therefore, either approach would be able to accommodate quantum noise, and the model observer performance at varying dose levels could be directly compared. This would also highlight the benefits of well-designed simulation studies as they are flexible, safe, and cost-efficient for evaluating many imaging system configurations.

3.5 Summary

We have presented a model observer that detects multiple signals and their locations. We have shown that the model observer using PLS channels and benchmark modified LG channels could achieve high performance with a small number of channels, but PLS channels significantly outperform LG channels when realistic signals and complex backgrounds were involved in the detection task. Rather than focusing solely on signals, it is essential for making multi-signal detection decisions to also consider local background statistics, signal prevalence, and the interactions among signal locations. We have also discussed that the optimal design of the model observer should be adjusted when the task of clinical interest changes, or when the resources such as computational power or training data are limited. Together, these results demonstrate that the model observer could be a useful tool in task-based assessment of image quality for cases with multiple abnormalities. For example, one of the practical applications of the model observers is with breast imaging modalities such as DBT. The model observer could be used to assess the potential of DBT to improve the detection of multiple breast tumors. It could help the optimization of DBT for diagnosis of MFMC breast cancer (e.g., design of system geometry, effective utilization of radiation dose).

In the next part of this dissertation, we present a series of studies with DBT to demonstrate the applications of our multi-signal model observer. In Chapter 4, we present background materials on MFMC breast cancer, and the potential of DBT in improving the detection of MFMC. The objective is to highlight the importance of including MFMC detection and diagnosis in the optimization of DBT. In Chapter 5, we present an extended 3D multi-lesion (ml) CHO model observer that detects multiple breast lesions from volumetric DBT imaging data. Chapter 6 presents a model observer study in which we employ the developed 3D ml-CHO to identify DBT system geometries that are most effective for the detection of MFMC.

Chapter 4

Digital Breast Tomosynthesis for Multifocal and Multicentric Breast Cancer

4.1 Introduction

In this chapter, we present background materials on MFMC breast cancer, and the potential of DBT in improving the detection of MFMC. The focus of the chapter is to highlight the significance of including MFMC detection in the optimization of DBT. This chapter is organized as follows: Section 4.2 briefly summarizes the clinical significance of MFMC (Section 4.2.1), and current roles of various medical imaging techniques in the detection of MFMC (Section 4.2.2); Section 4.3 introduces the fundamentals of DBT (Section 4.3.1) and its potential value in MFMC detection (Section 4.3.2).

4.2 Multifocal and Multicentric Breast Cancer

4.2.1 Clinical Significance of MFMC

Multifocality (MF) and multicentricity (MC) are commonly used in the initial assessment of the extent of breast cancer. MFMC is defined as multiple, simultaneous, physically separate, primary lesions when there are two or more foci of tumors present in the same breast [121, 45, 24]. Practically, MF breast cancer is defined as two or more tumor foci within a single breast quadrant while MC breast cancer is defined as two or more tumor foci within different quadrants of the same breast [64]. There is no exact radiological definition due to the difficulties in defining anatomically distinct borders between the breast quadrants and in estimating the precise distance between lesions. However, 5 *cm* is the most frequently used threshold of distance between tumors for distinguishing MF and MC (i.e., tumors are considered MF when the distance between tumors is less than or equal to 5 cm, and MC when the distance is more than 5 *cm*) [126]. The prevalence of MFMC has been reported to range from 9% to 75%, due to a lack of consistent assessment methods [167, 95, 39]. In the largest study of 8,935 breast cancer patients [155], the disease incidence was 20.8% (79.2% of the patients had a unifocal disease, 15.6% had MF tumors, and 5.2% had MC tumors).

MFMC can result from intra-mammary spread from a single primary tumor, or multiple synchronous primary tumors with inter-tumoral heterogeneity [41, 3, 91]. It is desirable to locate all of the lesions in the breast prior to the commencement of treatment because the objective is to remove all of them at the same time. Hence, a diagnosis of MFMC significantly impacts treatment planning. For instance, patients with MFMC are more likely to undergo mastectomy (i.e., surgery to remove the whole breast) than breast-conserving therapy (i.e., surgery to remove only the tumor and some surrounding tissue). The current TNM classification system [137] for staging MF/MC breast tumors, as recommended by the American Joint Committee on Cancer and the International Union against Cancer, is to use only the diameter of the largest focus [44]. As a result, the presence of additional tumor foci is not considered, which frequently underestimates the aggregate tumor burden [146]. Though the impact of MFMC on prognosis is still an ongoing controversy, MFMC is generally more aggressive than unifocal breast cancer, which usually leads to a higher recurrence rate, a higher metastatic rate, and an increase in lymph node involvement [39, 167, 95, 155, 3].

4.2.2 Imaging for the Detection of MFMC

Detecting multiple breast tumors with medical imaging is challenging as MFMC breast cancers are relatively uncommon, and the number and location of tumors is not known a *priori*. The prevalence of tumors also varies across breast regions (e.g., almost half of breast cancers are detected in the upper outer quadrant of the breast [80]). It is usually the case that the more conspicuous lesion is found first, and so finding the second lesion is more difficult as it will probably be smaller or in other ways less obvious. Cytology studies (e.g., [20, 43, 88]) have shown that the prevalence, and conspicuity of a particular type of targets may lead to missing subtle, smaller targets. Fleck et al. [53] showed in a multi-target search study that readers missed low-contrast targets more often when a high-contrast target was detected than on trials in which the low-contrast target was presented alone. Additional factors, such as bias about the expected number of tumors [53], satisfaction of search (i.e., detection of one target makes detection of subsequent targets less likely

[13, 125, 12]), may also impact the process of finding the second tumor.

The sensitivity of mammography for the detection of MFMC has been reported under 50%, ranges from 15% and 45% [60, 99, 93]. Breast ultrasound has been shown as a valuable complement imaging technique to mammography, and as a method of detecting additional tumors [14, 63, 108]. For example, the sensitivity for detecting MF lesions in [63] increased from 48% to 63% after adding sonograms to mammograms. In current clinical practice, women with breast cancers often undergo magnetic resonance imaging (MRI) to check for the presence of additional foci of disease that may be occult on mammography. Previous studies [64, 21, 126, 52] have shown that breast MRI has a superior sensitivity of $94 - 99\%$ in detecting multiple tumors. This advantage is even more evident for patients with dense breasts. However, apart from the drawbacks such as being time-consuming and expensive, the main disadvantage of breast MRI is its low specificity ($37\% - 86\%$ [60, 82]). Consequently, pre-operative MRI can lead to additional breast imaging, additional breast biopsies, and/or overtreatment, e.g., false-positive findings can result in women undergoing mastectomy rather than breast conserving therapy.

4.3 Digital Breast Tomosynthesis

4.3.1 Fundamentals of DBT

DBT is a new “3D mammography”, first approved by US Food and Drug Administration in 2011, formed by sweeping an x-ray beam in an arc over a limited angular range [127]. A series of low-dose projection images (usually 9-30) are acquired at different angles, and then reconstructed into a stack of 2D slices by shifting and adding the individual projections [127, 128]. As a result, a 3D picture of the breast tissue distribution is estimated. DBT is characterized with high in-plane spatial resolution (i.e., parallel to the detector), and a relatively lower depth resolution (i.e., perpendicular to the detector) [127]. DBT reduces tissue overlap relative to conventional mammography, and thus has the potential to improve breast cancer detection (i.e., higher sensitivity), decrease false alarms (i.e., higher specificity), and more accurately characterize lesions (e.g., [109, 59, 141, 22, 97]).

Different DBT system designs can yield different image characteristics such as spatial resolution (e.g., [66, 165]), lesion detectability (e.g., [119, 40, 32]), contrast-to-noise ratios (e.g., [84]) and observer preference (e.g., [57]). There are two design choices in image acquisition [127] that are of particular interest: 1) narrow- or wide-arc geometry (i.e., small or large angular span); and 2) a large or small increment of projection angles. An optimal DBT acquisition geometry can be obtained by balancing tradeoffs between and interaction effects of these two factors among many different geometry parameters. For examples, a larger angular span usually improves the depth perception of the breast, but it also decreases in-plane spatial resolution (e.g., [119, 86]). On the other hand, using a larger number of projection angles improves angular sampling. However, at a fixed total x-ray exposure and a fixed angular span, it also increases the amount of quantum noise in the reconstructed images [130]. Different projection-view distributions also affect the degrees of in-plane and inter-plane blurring in opposite ways [84]. In addition, increasing the angular span and the number of projection angles generally leads to a longer total scan time that may result in increased motion artifacts [57].

4.3.2 DBT for MFMC Detection

A few clinical studies have suggested that DBT has the potential to improve the detection of MFMC relative to MRI. For example, Wasserman et al. reported a case in which multiple foci were occult on MRI, but correctly identified on DBT [144]. These studies suggest that the potential sensitivity of DBT for MFMC may be equivalent to or higher than that of MRI, although no large comparison study has been published to date. DBT is expected to have a significantly higher specificity than MRI because the major cause of false positives on MRI (e.g., non-mass enhancement caused by factors such as mastopathic changes, fibrocystic changes due to hormonal stimulation [92]) is not relevant in x-ray imaging such as DBT. There are ongoing clinical trials (e.g., TOMODIAG in France [143]) that aim to investigate the potential of DBT in MFMC diagnosis. Overall, DBT may offer advantages over MRI in terms of fewer false-positive findings, lower cost, and better accessibility. In addition, DBT may be a better alternative than ultrasound when MRI is contraindicated; for example, Clauser et al. found a 23% increase in detecting additional lesions with second-look DBT relative to ultrasound [37]. However, prior efforts to optimize DBT system geometry for image quality only considered unifocal breast cancer scenarios (e.g., [127, 119, 66, 57, 130, 50]). DBT system geometries that yield images, that are informative for the task of detecting unifocal breast cancer, may not necessarily yield images that are ideally informative for the task of detecting MFMC. For example, a lesion seen in the specific sections of reconstructed images is in focus, but the out-of-focus blurring of the lesion in the depth direction may affect the detectability of another lesion, depending on its size and its distance from the in-focus lesion (e.g., [109, 89]). Lee et al. [81] showed that the detectability of small masses on DBT images varied across different breast regions in both in-plane and depth directions. Such factors have not been considered in prior optimization efforts with unifocal cases. Moreover, considering multiple tumors for DBT quality assessment raises new questions. For example, as tumors in MF cases are physically closer to each other than in MC cases, would higher in-plane resolution lead to higher lesion detectability? Would the anatomical variations in local breast regions as in MC cases cause less stable detection performance? These questions remain to be answered. Thus, optimizing DBT image quality for the detection of MFMC is important to enhance the potential clinical value of DBT.

In the next two chapters of this dissertation, we present a model observer study for optimizing DBT image quality for the detection of MFMC breast cancer. An extended 3D multi-lesion model observer that detects multiple breast lesions from volumetric DBT imaging data is developed in Chapter 5, and then employed to identify DBT system geometries that are most effective for MFMC detection in Chapter 6.

Chapter 5

3D Model Observer for Assessing DBT in Multi-lesion Detection

5.1 Introduction

Given the importance of including MFMC detection in the optimization of DBT as described in Chapter 4, it is essential to develop a practical model observer for detecting multiple lesions in volumetric image data[1,2]. This would also facilitate assessment and optimization of DBT image quality for improving human readers' interpretation of cases with potential MFMC lesions. In Chapter 3, we present our contribution of developing a 2D multi-lesion (ml) CHO for detecting multiple synthetic breast lesions in the presence of anatomical noise in synthetic 2D mammograms [152]. Unlike other existing multi-lesion model observers that rely heavily on unrealistic assumptions (e.g., stationary signal

---

[1] **Gezheng Wen**, Subok Park and Mia K. Markey, Digital breast tomosynthesis for detecting multifocal and multicentric breast cancer: influence of acquisition geometry on model observer performance in breast phantom images. *SPIE Medical Imaging*, **10136**: 101360V, 2017. (Gezheng Wen is the primary author who developed the methods, performed the analysis, and prepared the manuscript. Mia K. Markey and Subok Park helped with the study design and manuscript revisions.)

[2] **Gezheng Wen**, Mia K. Markey, Tamara Miner Haygood, and Subok Park. Model observer for assessing digital breast tomosynthesis for multi-lesion detection in the presence of anatomical noise. *in preparation* (Gezheng Wen is the primary author who developed the methods, performed the analysis, and prepared the manuscript. Mia K. Markey, Tamara Miner Haygood and Subok Park helped with the study design and manuscript revisions.)

[115], known background [7]), our 2D ml-CHO is adaptive to the characteristics of lesions and local background statistics. Novel implementations of both partial least squares (PLS) (Chapter 3.2.3) and LG channels (Chapter 3.2.4) have been developed to estimate different sets of more efficient channels that capture the correlations among signal locations, and help develop surrogate model observers for reliable observer performance levels and trends. Corresponding linear decision templates are employed to generate both image-level and location-specific scores on the presence of lesions (Chapter 3.2.5, Chapter 3.2.6). Under realistic imaging assumptions, the ml-CHO using the PLS channels and modified LG channels achieved high performance with a reasonably small number of channels (Chapter 3.3). We also show that the PLS channels outperformed the counterpart LG channels when realistic signal shapes and complex background statistics are involved. Hence, we argue that the 2D ml-CHO could be a reasonable starting point for us to modify for use in the assessment of DBT images.

In this chapter, we extend our previously developed 2D ml-CHO into a 3D volumetric ml-CHO that detects multiple lesions from volumetric DBT imaging data. Similar to the multi-signal detection task described in Chapter 3.2.1, the particular task of interest in the simulation study is to detect multiple synthetic lesions on 3D DBT images and make: 1) an image-level decision; and 2) a location-specific decision. In the preliminary version of this study presented and published in the proceedings of *SPIE Medical Imaging*

2017 [153], we presented the initial design of the 3D ml-CHO with the new implementations of 3D LG and PLS channels. The preliminary comparison results with multiple, identical spheres as simulated lesions showed that the 3D PLS channels outperformed the 3D LG channels chosen for the task. In the extended journal manuscript in preparation for a peer-reviewed journal [149], we further explore these questions with newly simulated, realistically shaped breast lesions. The evaluation results highlight the performance trends of the 3D ml-CHO in various clinical scenarios. Tests on factors that may contribute to the observed performance differences (e.g., choice of channels, local anatomical variations, and sample size) are also reported. In these two works, Gezheng Wen developed the methods, performed the analysis, and prepared the manuscript. Mia K. Markey, Tamara Miner Haygood, and Subok Park helped with the study designs and manuscript revisions.

This chapter is organized as follows: Section 5.2.1 describes the simulation of the MFMC image dataset, including information on the simulated DBT scanner in Section 5.2.1.1, digital breast phantoms in Section 5.2.1.2, and synthetic breast lesions in Section 5.2.1.3. In Section 5.2.2, we describe the model observer design and in Section 5.2.3, the strategies for evaluating the 3D ml-CHO.

5.2 Materials and Methods

5.2.1 Image Dataset

5.2.1.1 Simulated DBT scanner

The simulated DBT system has an ideal photon-counting flat panel detector and a point x-ray source that moves along an arc trajectory in a step-and-shoot mode [160]. The distance between the x-ray source and the detector was 65 *cm* and the rotation center of the x-ray tube was 6 *cm* above the detector. When the x-ray source rotates, the detector moves along the x-y plane. The x-ray scatter in photon transport was ignored, and only Poisson noise was considered. A 20 *keV* monoenergetic x-ray spectrum with a fixed total exposure at all scan settings was simulated. The total amount of exposure at the detector was approximately 30 mR, which was matched with the exposure level measured with real DBT scanners in similar scan settings (e.g., a mean glandular dose of 1.6 *mGy* to a 4 *cm* thick breast as measured in [165]). For imaging acquisition, we used a representative DBT system geometry that covered an angular span of 24 degrees, and obtained 9 projection views per scan. This geometry of image acquisition (not including the detector configuration) was similar to the GE SenoClair DBT system (General Electric Healthcare, Milwaukee, WI) that uses 25-degree angular span and 9 projection views [127]. Reconstruction slice thickness was $1mm$ and the in-plane resolution was 125 μm. We used the conventional filtered back-projection (FBP) with a Hann filter for reconstruction.

5.2.1.2 Structured image background

Digital breast phantoms mimic breast anatomy (e.g., shape of breast, and texture within each breast), and random structural variations between patients as needed to evaluate task-based image quality (e.g., [6, 65, 70]). We used Bakic et al.'s 3D computational breast phantom [6], which uses a region growing method to simulate basic breast anatomical components including skin, adipose tissue, glandular tissue, and Coopers ligaments. 5000 lesion-free phantoms were generated [153]. The breast phantom of size $204.5 \times 65 \times 51.5$ mm^3 with the isotropic voxel size 0.5 mm was set with a volume of 450 ml (i.e., cup size B), compressed thickness of 51.5 mm, and volumetric glandular fraction of 25%. To match the phantom voxel size with the finer detector element size of 125 μm, a 3D linear interpolation was applied to up-sample the phantoms by a factor of 4 following the approach in [119, 160, 55]. To further improve the realism of the phantoms by simulating small-scale, high-frequency anatomical structures, we followed the method that was generalized by the method in [79]. The volumes of random 3D power-law noise of coefficient 3.0 were then binarized by global thresholding (i.e., voxels below and above the chosen threshold of 0.45 were set to 0 and 1, respectively), and added to a limited fraction of dense breast regions (i.e., 0.5 of the glandular fraction as labeled for the phantom components) [79].

5.2.1.3 Synthetic breast lesions

We simulated synthetic 3D breast lesions by analytically defining the shape of each lesion using a set of basis vectors in spherical coordinates [138]. Adapted from the publicly available lesion database of [138], 32 combinations of the radius R_b and the radial direction (θ_b, ϕ_b) were used to cover a range of typical lesion shapes (e.g., oval, lobulated, irregularly shaped with spikes). The voxelization process included two steps: 1) Compute the radial direction (θ, ϕ) for each voxel in a 3D Cartesian grid (i.e., (x, y, z) with the origin at the center of the lesion); and 2) Perform an interpolation on R_b in spherical domain to find the radius $R(\theta, \phi)$ of the lesion in the given direction (θ, ϕ). The resulting interpolated surface was then used as the enclosing boundary of a solid 3D volume (i.e., a lesion). As the voxel spacing for these mathematically defined lesions could be arbitrarily small, we set the voxel size of the lesions to be 125 μm (i.e., the same voxel size of the breast phantom). The lesion volumes had an approximate diameter of 6 $mm \pm 1$ mm, and an approximate volume of 80 $mm^3 \pm 30$ mm^3. In addition, we simulated spiculated structures using a modified version of the iterative fractal branching algorithm in Ref. [135]. That is, a random set of spicules was created by recursively creating continuing branches in the form of conical frustums (i.e., each of the branches was characterized by a starting and an ending 3D location, a starting and an ending radius, and length). Starting with an initial branch length of 0.08 mm and radius of 0.04 mm, the branch length and radius reduced at each iteration, and the direction of extension changed slightly. To simulate the

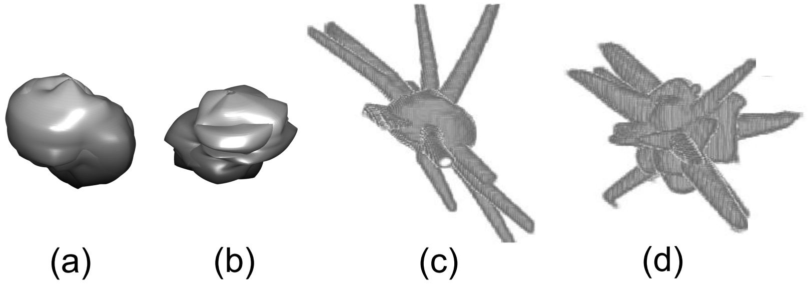


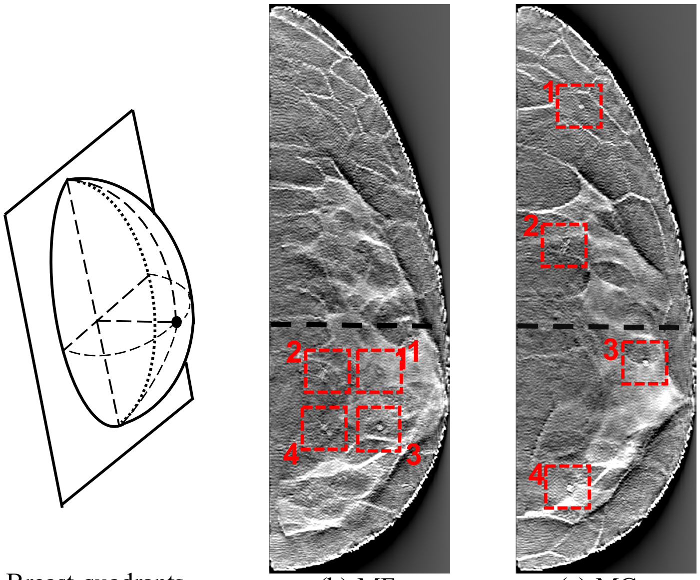
Figure 5.1: The four types of simulated breast lesions used in the evaluation study (i.e., the lesions to be detected by a model observer): (a)-(b): mathematically defined, non-spiculated lesions, (c)-(d) spiculated lesions after embedding spiculated structures to the well-circumscribed lesions

variations across different spiculated lesions, all the growing rules included a set of random variables, which followed the pre-defined normal distributions as follows: 1). the decrease ratio of branch length (i.e., fraction of the previous length) $\sim N(0.65, 0.1)$; 2). the decrease ratio of radius (fraction of the previous radius) $\sim N(0.5, 0.2)$; 3). the angle change of branch extension (in degrees) $\sim N(10, 5)$. The four types of lesions used in the evaluation study (i.e., the lesions to be detected by a model observer) are shown in Figure 5.1.

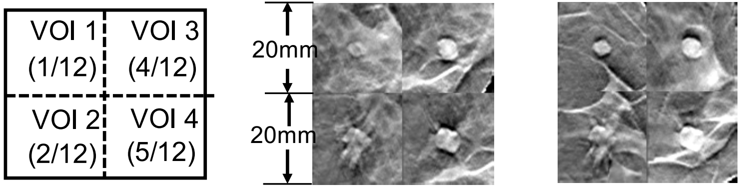
Lesion-present breast phantoms were generated by embedding a random number of breast lesions in the phantoms to mimic MF and MC cases (i.e., replacing the affected voxels of phantoms by the voxels of the lesions). We divided the breast into four quadrants (Figure 5.2(a)), and there were four possible lesion locations across the breast (Figure 5.2(b) for MF cases and Figure 5.2(c) for MC cases) . The lesion locations were sampled from a prevalence distribution $\{\frac{1}{12}, \frac{2}{12}, \frac{4}{12}, \frac{5}{12}\}$ (Figure 5.2(d)). The number of lesions was sampled from a truncated Poisson distribution with a mean of 1. All the assumptions model the typical incidence and distribution of breast lesions detected in different breast regions [80, 100, 15, 131], and how radiologists usually report the location of a detected lesion.

(i). For MF cases, adjacent lesions were separated from each other by 30 mm (Figure 5.2(b)). The separation distance was selected such that all the lesions appeared in the glandular region of one breast quadrant, and there was also some correlation between the local volumes of interest (VOI) containing lesions. Four local VOIs of in-plane size 20 $mm \times 20$ mm at each lesion location (either lesion-present or lesion-absent) were extracted, and then combined into a large 2×2 VOI, denoted as $\mathrm{VOI}_{\mathrm{MF}}$ (Figure 5.2(e)).

(ii). For MC cases, lesions were located across different breast quadrants (Figure 5.2(c)), which resulted in different local background statistics [110, 120, 87]. Similarly, four local VOIs at each lesion location were extracted, and combined into a large 2×2 VOI, denoted as $\mathrm{VOI}_{\mathrm{MC}}$ (Figure 5.2(f)). This was done to simply the problem of detecting MC cases, but the $\mathrm{VOI}_{\mathrm{MC}}$ still captures some of the potential spatial correlations between the four local VOIs.



(a) Breast quadrants (b) MF (c) MC



(d) local VOIs (e) VOI$_{\text{MF}}$ (f) VOI$_{\text{MC}}$

Figure 5.2: Example DBT reconstructed slices, and the corresponding whole VOI slices of two MFMC cases. (a) shows the 3D schematic of four breast quadrants, in which the *thin dashed* lines define the anatomical boundaries for separating the breast into four quadrants. The two lesion-present phantoms in (b) and (c) were scanned. (b) and (e) show an example MF case with 4 lesions that are present in the same breast quadrant. (c) and (f) show an example MC case with 4 lesions that are present in different breast quadrants. The *red dashed* boxes in (b) and (c) indicate the extracted local VOIs centered at the four possible lesion locations, where the *red* digits (i.e., 1-4) indicated the local VOI indices. The *black dashed* lines are the quadrant boundaries. (d) shows the schematic of local VOI locations in the whole VOI (i.e., VOI$_{\text{MF}}$ or VOI$_{\text{MC}}$), and the corresponding lesion prevalence. In (e) and (f), all four lesions are contrast-enhanced for better visualization: the two lesions on the top (i.e., local VOI 1 and 3) are non-spiculated lesions, while the two lesions on the bottom (i.e., local VOI 2 and 4) are spiculated lesions. Lesion shapes in the local VOIs of all the simulated MFMC cases were of the same order as in (e) and (f).

5.2.2 3D Model Observer for Multi-lesion Detection Task

We extended our previously developed 2D model observer and channel approaches (Chapter 3) [152] to a volumetric observer, the 3D ml-CHO, for detecting multiple lesions on DBT reconstructed image slices [153, 149]. The main idea is to consider 3D DBT reconstructed image data (varying the number of slices N_s included the VOI) as a single object to be examined, and that detection decisions regarding the presence of lesions are made upon the overall information from both individual slices and their correlations [101, 114]. There are three main steps for the 3D ml-CHO to generate the detection variables: 1) concatenate the stack of relevant reconstructed image slices together to represent the 3D DBT image data $\mathbf{X}$; 2) compute the location-level 3D channels $\mathbf{T}_{\text{loc}}$ [151, 152]; and 3) use the estimated $\mathbf{T}_{\text{loc}}$ channels in the 3D ml-CHO for generating decision variables [114].

As stated in Table 5.1, each column of $\mathbf{X}$ is of length $P \times N_s$, and it represents the concatenated image data from N_s slices. Specifically, the j^{th} column of $\mathbf{X}$, $\mathbf{X}_j$, is written as:

$$\mathbf{X}_j = \{\mathbf{f}_m\}_{m=s_1}^{s2}, \{\mathbf{f}_m\} = \{f_{m,i}\}_{i=1}^{P}$$

where $\{\mathbf{f}_m\}$ is the m^{th} reconstructed image slice of P image pixels, and m ranging from slice s_1 to slice s_2 indicates which subset of the slices is included for the given detection task.

Each 3D $\mathbf{T}_{\text{loc}}$ channel, as a stack of N_s 2D channels, captures correlations within and between the slices. We computed both 3D PLS and modified 3D LG channels for comparisons as follows. First, the PLS channels were estimated directly from training images (i.e., $\mathbf{T}_{\text{loc}} = \mathbf{PLS}(\mathbf{X}, \mathbf{Y}^{\text{L}})$, where $\mathbf{PLS}$ is the PLS channel generating function (Chapter 3.2.3) that maximizes the covariance between $\mathbf{X}$ and the location-specific truth labels $\mathbf{Y}^{\text{L}}$ [152]). By design, the PLS channels were naturally adaptive to complex anatomical backgrounds and signal statistics. Second, following the approach we developed in Chapter 3.2.7.2, the counterpart LG channels were analytically derived based on the empirical occurrence frequency p of lesion-presence patterns in the training set (i.e., how many and where lesions appear at the same time: 4 scenarios with one lesion, 6 scenarios with two lesions, 4 scenarios with three lesions, 1 scenario with all four lesions) [152]. For each scenario with $p > 0.025$ (i.e., to mimic a 95% confidence), we added 3D LG channels centered at each of the lesion-present locations. The Gaussian widths of those 3D LG channels along the x, y, and z directions were fixed at 6 mm, mimicking the size of the simulated lesions. Starting with 0^{th}-order LG channels, we looped over all scenarios, in the descending order of p, before repeating the steps with 1^{st}-order LG channels. An orthonormal set of channels was extracted as the final $\mathbf{T}_{\text{loc}}$ by orthonormalization. One advantage of LG channels is that they are based on an analytical expression, so the quality of LG channels does not depend on the number of available training samples [152].

The linear discriminant template $\mathbf{w}^{\mathrm{L}}_{\mathrm{loc}}$, as a function of $\mathbf{T}_{\mathrm{loc}}$, $\mathbf{X}$, and $\mathbf{Y}^{\mathrm{L}}$, were trained for computing decision variables $\mathbf{S}^{\mathrm{L}}_{\mathrm{loc}}$:

$$\begin{aligned}\mathbf{w}^{\mathrm{L}}_{\mathrm{loc}} &= \mathbf{K}_{\mathbf{V}}^{-1}\bar{\mathbf{V}} \\ \mathbf{S}^{\mathrm{L}}_{\mathrm{loc}} &= (\mathbf{w}^{\mathrm{L}}_{\mathrm{loc}})^{T}(\mathbf{T}^{T}_{\mathrm{loc}}\mathbf{X})\end{aligned}$$

where $\mathbf{V} = \mathbf{T}^{T}_{\mathrm{loc}}\mathbf{X}$ (^{T} indicates transpose) is the set of channelized images that are classified into $\mathbf{V}_1$ and $\mathbf{V}_0$ based on $\mathbf{Y}^{\mathrm{L}}$ (i.e., $\mathbf{V}_i(j)$ are the columns that $\mathbf{Y}^{\mathrm{L}}(j)$ is i, where $i = 0$ or 1, $j = 1, 2, .., N_l$), $\bar{\mathbf{V}} = \bar{\mathbf{V}}_1 - \bar{\mathbf{V}}_0$ is the mean channelized signal(s), and $\mathbf{K}_{\mathbf{V}} = \frac{1}{2}(\mathbf{K}_{\mathbf{V}_1} + \mathbf{K}_{\mathbf{V}_0})$ is the mean covariance matrix of $\mathbf{V}_1$ and $\mathbf{V}_0$ [61].

The higher the score is, the more likely the lesion is present. We note that $\mathbf{S}^{\mathrm{L}}_{\mathrm{loc}}$ contained location-specific scores that were used for making both image-level and location-specific decisions. For image-level decisions, a second-stage linear decision template was used to convert $\mathbf{S}^{\mathrm{L}}_{\mathrm{loc}}$ into scalar image-level scores [152, 114]. Table 5.1 summarizes the notation used in the 3D ml-CHO.

5.2.3 Evaluation

Similar to the evaluation strategies in Chapter 3.2.7.1 for 2D ml-CHO, we used two FOMs to measure observer performance: 1). For image-level decision variables, a task SNR was computed to quantify the binary classification performance [61]. 2). For location-specific decision variables, AFROC curves were plotted, and $\mathrm{AUC}_{\mathrm{AFROC}}$ over $[0, 1]$ was used as the other FOM.

Table 5.1: Notation of image data and response for 3D ml-CHO

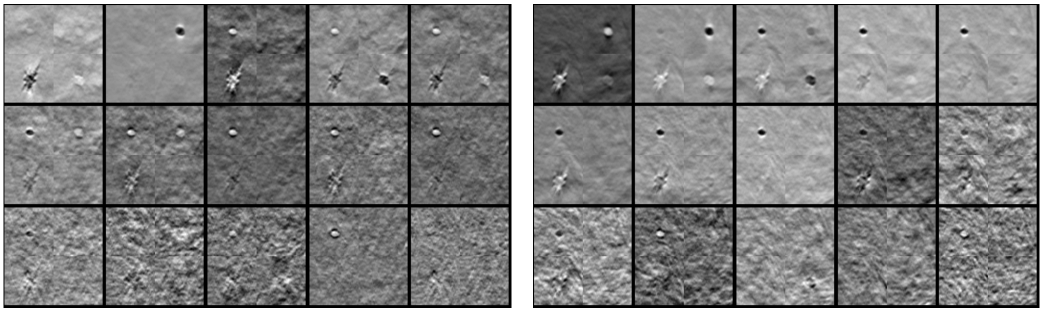
Notation	**Dimension**	**Descriptions**
N	scalar	number of images
P	scalar	number of pixels in each image
N_l	scalar	number of possible lesion locations
N_s	scalar	number of reconstructed DBT slices
N_c	scalar	number of 3D channels
$\mathbf{X}$	$(P \times N_s) \times N$	image data
$\mathbf{Y}^{\mathrm{L}}$	$N_l \times N$	location-specific truth labels
$\mathbf{T}_{\mathrm{loc}}$	$(P \times N_s) \times N_c$	location-level 3D channels
$\mathbf{w}_{\mathrm{loc}}^{\mathrm{L}}$	$N_c \times N_l$	linear decision template
$\mathbf{S}_{\mathrm{loc}}^{\mathrm{L}}$	$N_l \times N$	location-specific decision variable

For each run of the evaluation experiment, we first embedded a random number of lesions to be detected (i.e., ranging from 0 to 4 based on the Poisson distribution of mean 1) into the 5000 lesion-free structured breast phantoms. Then we randomly divided these simulated MFMC cases into three non-overlapping phantom groups: 2000 cases for estimating the 3D PLS channels $\mathbf{T}_{\mathrm{loc}}$, 2000 for training the linear discriminant template $\mathbf{w}_{\mathrm{loc}}^{\mathrm{L}}$, and the other 1000 cases for testing. Stratified random sampling [38] was used to assure that the prevalence of lesion-presence scenarios in each of the three groups were approximately the same (i.e., 36.8% of lesion-absent cases, 36.8% of one-lesion cases, 18.4% of two-lesion cases, 6.1% of three-lesion cases and 1.9% of four-lesion cases). This was done by taking a random subset of cases from each of the lesion-presence scenarios in a number proportional to its exact occurrence frequency. These subsets were then pooled to form the three groups accordingly. To investigate the impact of N_s on MFMC detection, the 3D ml-CHO model was trained independently with six different sets of $\mathbf{X}$ (i.e., $\mathbf{X}$ with $N_s = 1, 3, 5, 7, 11$, or 15), resulting in six different model observers. The experiment was repeated for 20 runs, where each run used a different division of the three groups. The average and two standard deviations of the task SNR and $\text{AUC}_{\text{AFROC}}$ were reported to identify performance trends.

5.3 Results

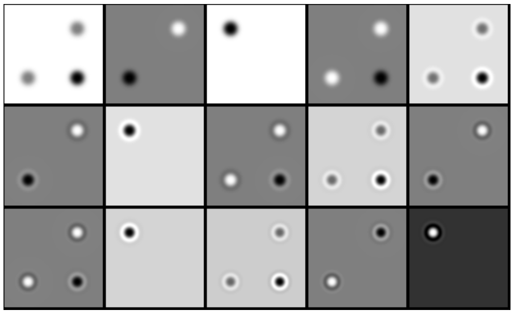
5.3.1 3D $\mathbf{T}_{\text{loc}}$ Channels

Figure 5.3 shows example 3D $\mathbf{T}_{\text{loc}}$ PLS channels estimated for MF cases (Figure 5.3(a)), and for MC cases (Figure 5.3(b)). In this example, each of the VOI_{MF} and VOI_{MC} contained 5 slices (i.e., $N_s = 5$), but only the first 15 channels (i.e., $N_c = 15$) for the central slice of the whole VOI are plotted. For both MF and MC cases, the following three observations could be made. First, $\mathbf{T}_{\text{loc}}$ captured the characteristics of both non-spiculated and spiculated lesions (e.g., the spiculated structures of the spiculated lesion in local VOI 2 were clearly shown around the central volume of the lesion). Second, lesion contrasts and lesion-presence patterns (i.e., which of the lesions were more likely to be present together in the same case) in the $\mathbf{T}_{\text{loc}}$ channels reflected lesion prevalence at different lesion locations, and the interactions between lesion locations. Third, $\mathbf{T}_{\text{loc}}$ were adaptive to local background statistics. Hence, $\mathbf{T}_{\text{loc}}$ were used as the common set of channels to generate both image-level and location-specific decisions (Chapter 3.2.3)[152]. Figure 5.3(c) shows example



(a) PLS: MF

(b) PLS: MC



(c) LG: MF

Figure 5.3: (a)-(c) The first 15 columns of example 3D $\mathbf{T}_{\text{loc}}$ channels: 1st row shows the 1st (*left*) to 5th (*right*) channel, 2nd row shows the 6th (*left*) to 10th (*right*) channel and etc. (a) PLS $\mathbf{T}_{\text{loc}}$ for MF cases; (b) PLS $\mathbf{T}_{\text{loc}}$ for MC cases; (c) LG $\mathbf{T}_{\text{loc}}$ for MF cases. PLS $\mathbf{T}_{\text{loc}}$ in (a) and (b) captured leison characteristics. Lesion contrasts and lesion-presence patterns in the PLS channels also reflected lesion prevalence at different locations, and the interactions between locations. LG $\mathbf{T}_{\text{loc}}$ in (c) reflected the possibility of lesion-presence patterns.

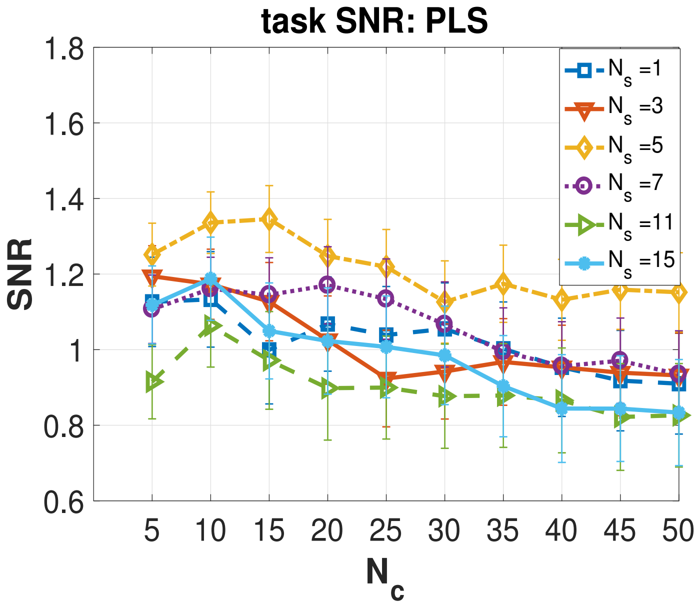


Figure 5.4: Task SNR trends for the detection of MF lesions as a function of N_c for six different values of N_s using the 3D PLS channels. The SNR was the highest when $N_s = 5$ ($yellow$). Increasing N_c did not necessarily lead to a higher SNR.

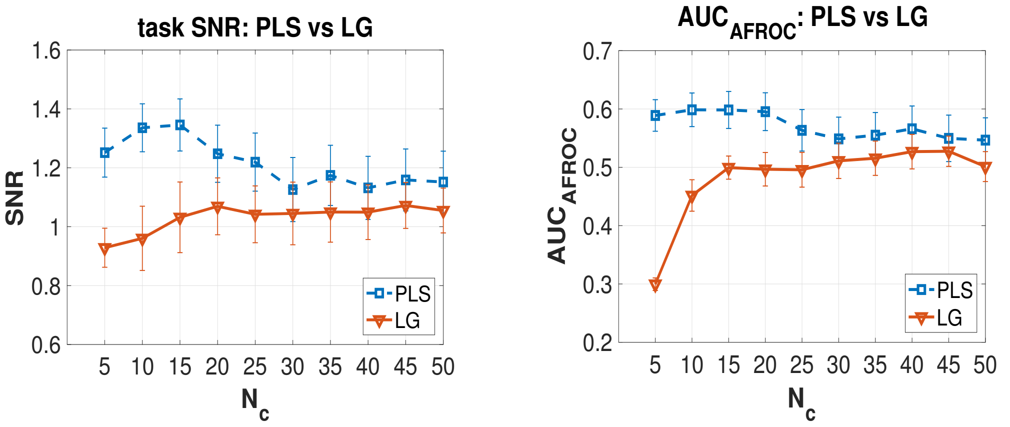
3D $\mathbf{T}_{\text{loc}}$ LG channels for MF cases (i.e., counterpart to the PLS channels in Figure 5.3(a)). These analytically derived symmetric LG channels reflected the possibility of lesion-presence patterns (e.g., the scenarios with three present lesions were the most possible empirically). However, they did not capture as much local background, or lesion characteristics that are essential for the given detection task as the PLS channels did given the same number of channels.

5.3.2 FOMs versus N_s and N_c

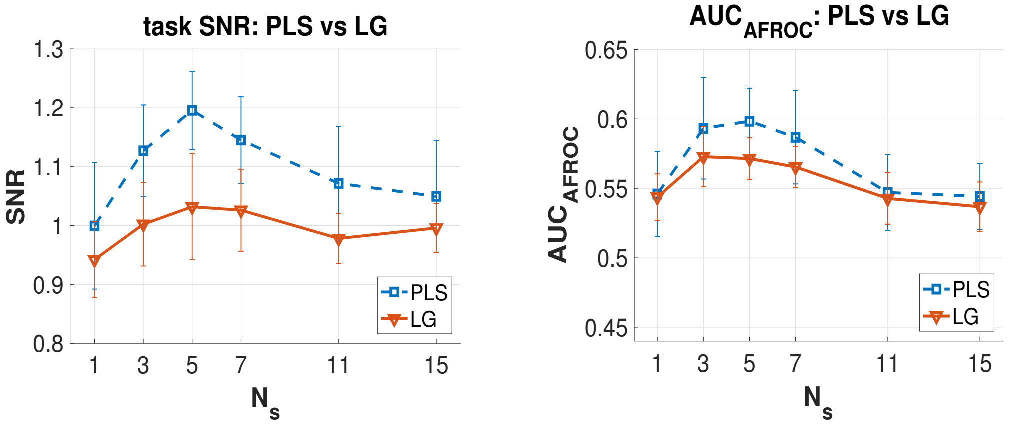
We investigated how the detection performance of the 3D ml-CHO changed as the numbers of channels and reconstructed slices, N_c and N_s, were varied. Figure 5.4 shows the estimated values of mean task SNR as a function of N_c for the case of MF lesions. The results with the 3D PLS channels and six different N_s values were reported. When $N_s = 5$ (the *yellow* curve), the SNR was the highest among all the N_s cases compared. This reveals that including additional slices into the 3D ml-CHO model may not always improve lesion-detection performance. Given the mechanism of the 3D ml-CHO model that considers the stack of slices as a single object to be examined, the added slices may play as irrelevant noise, resulting in less discriminating decision variables for the given detection task. Across all N_s cases considered, the SNR estimates tend to peak around $N_c = 10 \sim 20$ and decrease/plateau as N_c is increased to 50. This indicates that increasing N_c beyond 20 did not necessarily lead to a higher SNR. This was likely due to the limited number of training images available for accurately and stably estimating the 3D PLS channels and decision templates $\mathbf{w}$ when N_c was relatively high. For example, if 40 3D PLS channels were applied to a VOI of 5 slices (i.e., $N_s = 5$, $N_c = 40$), the total number of elements in the channelized data was 40. As a result, the size of channelized covariance matrix was 40×40, which was approximately the same as the training sample size. More investigations of the effect of training sample size on the 3D ml-CHO performance trends for the same detection task are reported in Section 5.3.5. Similar overall performance trends were observed with: 1) image-level detection performance (task SNR) for the MC cases; 2) location-specific detection performance ($\text{AUC}_{\text{AFROC}}$) for both MF and MC cases; and 3) the 3D ml-CHO with the LG channels. Thus, for the rest of the Results section, we will report the DBT evaluation results with $N_s = 5$ and/or $N_c = 15$, if N_s and/or N_c need be fixed for comparison.

5.3.3 PLS Channels versus LG Channels

Figure 5.5 presents the comparison between the ml-CHO with 3D PLS channels and that with 3D LG channels. The estimated values of mean task SNR and $\text{AUC}_{\text{AFROC}}$ for MF cases were reported, and similar trends were observed for MC cases. Figure 5.5(a) and 5.5(b) shows the task SNR and $\text{AUC}_{\text{AFROC}}$ with respect to varying N_c with the fixed N_s of 5, and Figure 5.5(c) and (d) shows the FOMs with respect to varying N_s with fixed N_c of 15. It is evident that the ml-CHO with the PLS channels (*blue*) outperformed that with the LG channels (*red*) in all four scenarios. This advantage of PLS channels could be explained by the fact that PLS channels were estimated directly from the images, and thus they were naturally adaptive to signal characteristics and local background statistics (Figure 5.3(a) and (b)). In contrast, analytically derived LG channels (Figure 5.3(c)) only reflected the possibility of signal-presence patterns (e.g., the scenarios with three present lesions were the most possible empirically). We set the widths of LG channels to mimic the size of the lesions, but prior studies [160, 54, 103] have shown that the optimal LG widths are affected by the background complexity. Potentially insufficient tuning of the LG channels, especially for the 3D DBT image data in this study, may have degraded the performance of ml-CHO with LG channels. We also recognize that rotationally symmetric LG channels are most effective with ro-



(a) task SNR versus N_c: $N_s = 5$ (b) $\text{AUC}_{\text{AFROC}}$ versus N_c: $N_s = 5$

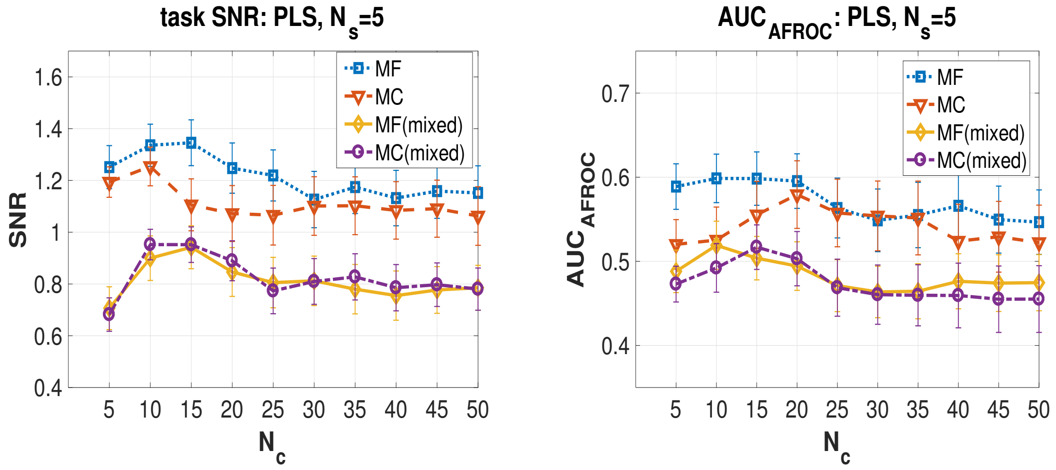


(c) task SNR versus N_s: $N_c = 15$ (d) $\text{AUC}_{\text{AFROC}}$ versus N_s: $N_c = 15$

Figure 5.5: Example task SNR and $\text{AUC}_{\text{AFROC}}$ achieved by the 3D ml-CHO with 3D PLS channels in Figure 5.3(a) and 3D LG channels in Figure 5.3(c): a) task SNR with respect to varying N_c with fixed N_s of 5; (b) $\text{AUC}_{\text{AFROC}}$ with respect to varying N_c with fixed N_s of 5; (c) task SNR with respect to varying N_s with fixed N_c of 15; (d) $\text{AUC}_{\text{AFROC}}$ with respect to varying N_s with fixed N_c of 15. The error bars indicated two standard deviation of the FOM above or below the mean FOM. It shows that the PLS channels (*blue*) outperformed the LG channels (*red*) in all four scenarios.

tationally symmetric signals (e.g., spheres) [61, 81, 101]; hence, there would be a substantial decrease in detection performance for complex detection tasks.

We observe that for both PLS and LG channels, increasing N_c (≥ 10) did not necessarily improve the mean task SNR or $\mathrm{AUC}_{\mathrm{AFROC}}$. As briefly discussed in Section 5.3.2, this could be due to the limited number of training images for accurately and stably estimating decision templates w when N_c was relatively high. The estimation of the PLS channels may be limited by the number of training images as well. This led to one of the key disadvantages of PLS channels to LG channels as the quality of analytically derived 3D LG channels was independent of the number of channel training images. The performance of the ml-CHO with both 3D LG and PLS channels may be affected by the particular set of training images, as 1) the LG channels were generated based on the empirical occurrence frequency of lesion-presence patterns (Section 5.2.2), and 2) the PLS channels are generated from the training images themselves. In addition, it is possible that beyond a certain range of N_c, adding more channels would not capture any additional discriminant information for the detection task, while hurting the stability of estimation of the high-dimensional covariance in observer calculation. Similarly, increasing N_s did not necessarily increase the FOMs. One of the possible explanations was that including additional slices may add more of distraction rather than useful discriminant information, as lesions may not appear in all of the used slices. Given that 1) the trends of the task SNR as a function of N_c are different be-



(a) task SNR

(b) $\mathrm{AUC}_{\mathrm{AFROC}}$

Figure 5.6: FOM comparisons between the ml-CHOs trained with (*yellow*, *purple*) or without (*blue*, *red*) mixing the four local VOIs. 5 reconstructed slices with 3D PLS channels were used by each of the ml-CHOs (i.e., $N_s = 5$). The results show that mixing the local VOIs into one ensemble for ml-CHO training significantly degraded the detection performance (*yellow* lower than *blue*, *purple* lower than *red*). The ml-CHO performance for MF (*yellow*) and MC cases (*purple*) were approximately identical.

tween PLS channels and LG channel, and 2) the task SNR values as a function of N_s are higher with PLS channels than those with LG channel, we will report evaluation results only with 3D PLS channels in the rest of the Results section.

5.3.4 Impact of Locally Varying Anatomical Background

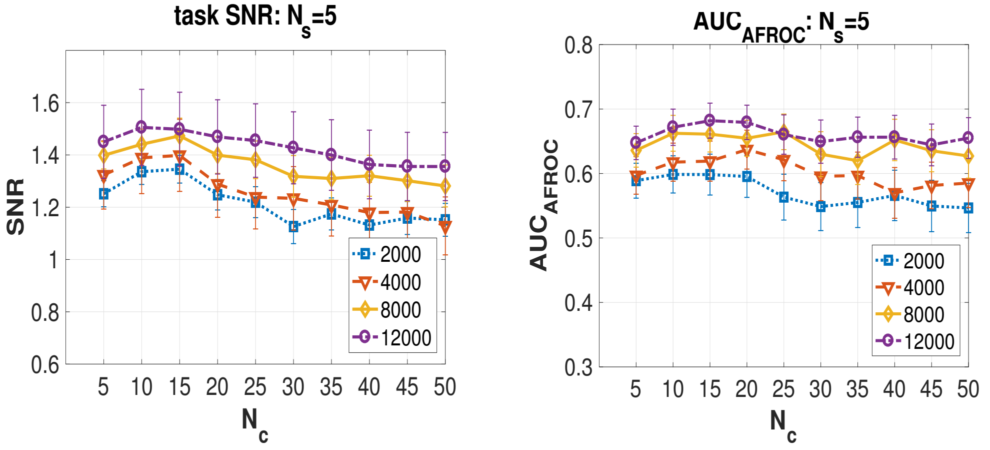
The local VOIs in the MC cases were located in different breast quadrants, and hence local background statistics of the local VOIs may be significantly different. In contrast, for the local VOIs in MF cases, they were close to each other, so spatial distributions of breast tissues within these local breast regions were expected to be less dissimilar than in MC cases. To study the impact of locally varying anatomical backgrounds on the detection of MF/MC lesions, we compared the performance trends of 3D ml-CHO when the four local VOIs were 1) handled separately as described in Section 5.2.2, and 2) mixed into one single ensemble for training the 3D ml-CHO (i.e., the ml-CHO assumed the local VOIs to be not statistically different from each other). The second scenario was done by assigning random indexing on the local VOIs so that there was no bias toward which local VOI was coming from which local region of the breast. As a result, the 3D PLS channels and CHO template $\mathbf{w}$ were in fact estimated using the four local VOIs (i.e., the ones forming $\mathrm{VOI}_{\mathrm{MF}}$ or $\mathrm{VOI}_{\mathrm{MC}}$) extracted from one identical distribution although the same DBT image set as in Section 5.2.1 was used. In each of the subfigures in Figure 5.6, the curves for MF (*blue*) and MC cases (*red*) are the ml-CHO performance when the local VOIs were handled separately, and the curves for MF(mixed) (*yellow*) and MC(mixed) (*purple*) are the counterpart results when mixing the four local VOIs. The horizontal axis of the plots is the number of PLS channels N_c used in the ml-CHO, and N_s was set to 5. It can be observed that:

(i). For both the task SNR and $\mathrm{AUC}_{\mathrm{AFROC}}$, mixing the local VOIs for training the ml-CHO significantly degraded the detection performance of the 3D ml-CHO (*yellow* lower than *blue*, *purple* lower than *red*). This agrees with our prior studies in Chapter 3.3.2 that incorporating correlations among local anatomical background into the ml-CHO design is beneficial for multi-lesion detection tasks (e.g., the presence of one lesion may alter the detection at other regions due to potential concurrence) [152]. Moreover, the performance trends of the ml-CHOs as a function of N_c were different between the mixed and non-mixed VOI cases, more pronounced in the case of MF lesions. For example, in Figure 5.6(a), the task SNR for MF cases with the mixed VOIs (*yellow*) was highest when $N_c = 15$, which was not true for the SNR without mixing the VOIs (*red*).

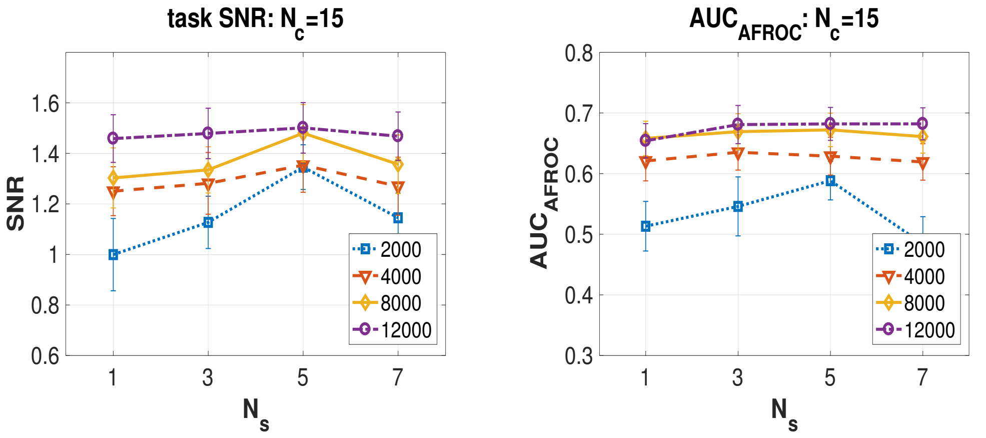
(ii). Both performance levels and trends of the ml-CHOs were almost identical between the MF (*yellow*) and MC cases (*purple*) when the four local VOIs were mixed together. The negligible gaps in the FOMs could be due to the finite sample sizes that may lead to slightly different sample statistics for the MF cases and MC cases.

5.3.5 Sample Size Consideration

We have shown in Section 5.3.2 that the 3D ml-CHO was able to make accurate image-level and location-specific detection decisions with a small number of channels. For example, it achieved a high mean task SNR of 1.37 with N_s=5 at N_c =10 (Figure 5.4(a)), which was significantly higher than the SNR of 0.95 when the local VOIs were mixed (Figure 5.6(a)). However, increasing N_c or N_s did not always lead to a higher FOM, especially for large N_s. To understand how the limited number of phantoms used in



(a) task SNR versus N_c: $N_s = 5$ (b) $\text{AUC}_{\text{AFROC}}$ versus N_c: $N_s = 5$



(c) task SNR versus N_s: $N_c = 15$ (d) $\text{AUC}_{\text{AFROC}}$ versus N_s: $N_c = 15$

Figure 5.7: Observer performance trends versus N_s or N_c with four different training sample sizes for the MF cases. Plots (a) and (b) are the task SNR and $\text{AUC}_{\text{AFROC}}$ trends as a function of N_c when $N_s = 5$. Plots (c) and (d) are the task SNR and $\text{AUC}_{\text{AFROC}}$ trends as a function of N_s when $N_c = 15$. The trends of these FOMs as a function of N_c or N_s were similar across the different training sample sizes. The plots indicate that 1) increasing N_c or N_s did not always result in a higher task SNR or $\text{AUC}_{\text{AFROC}}$; and 2) overall the highest task SNR and $\text{AUC}_{\text{AFROC}}$ were achieved when $N_c = 15$ and $N_s = 5$.

this study might have impacted the performance trends, we assessed model-observer performance with various sizes of training sets. We fixed the number of testing cases at 1000, and varied the number of training cases to one of $\{2000, 4000, 6000, 12000\}$. One half of the training cases were used to estimate the 3D PLS channels, and the other half were used to train the decision templates. As there were only 5000 independent breast phantoms, we sampled the phantoms with replacements to achieve a larger sample size. Similarly, the model observer experiment with each setting was repeated for 20 runs in order to estimate the average and variance of model-observer performance.

As an illustration given in Figure 5.7, we focused on the MF cases, but similar trends were seen for the MC cases as well. Despite the notable differences in the absolute performance, the trends of task SNR and $\text{AUC}_{\text{AFROC}}$ as a function of the number of 3D PLS channels N_c (Figure 5.7(a), (b)) and as a function of the number of reconstructed slices N_s (Figure 5.7(c), (d)) were similar across the different numbers of training phantom cases considered. Increasing N_c or N_s did not always result in a higher SNR or a higher $\text{AUC}_{\text{AFROC}}$. Overall, the highest task SNR and $\text{AUC}_{\text{AFROC}}$ for all four training sample sizes were achieved when N_c was approximately 15, and when $N_s = 5$ among the N_s values considered, respectively. This was consistent with the performance trends shown in Figure 5.4. Hence, with the limited number of breast phantoms, the 3D ml-CHO with 3D PLS channels may still efficiently capture useful image discriminant information for predicting the trend of accuracy in making detection decisions, even if the quality of the channels and decision templates may not be as good as they could possibly be with a larger training sample size. Changing the number of training cases would not significantly alter the qualitative conclusion of the study.

5.4 Discussion

We present a 3D volumetric ml-CHO model to detect multiple lesions in simulated DBT reconstructed image slices. The 3D PLS channels were estimated directly from the DBT image data, and they were adapted to variations in breast lesions and anatomical backgrounds. They also captured the spatial correlations within and across the DBT slices. The 3D LG channels were analytically derived, and their quality does not depend on the number of channel training images. We have shown that the 3D ml-CHO with the PLS channels was able to make creditably accurate image-level and location-specific detection decisions with a small number of channels N_c (e.g., $N_c = 10$), and it outperformed the counterpart ml-CHO with 3D LG channels. However, increasing N_c or N_s did not always lead to a higher mean task SNR or $\text{AUC}_{\text{AFROC}}$, especially for large N_s. It is possible that the trends as a function of N_c or N_s would be clearer if a larger number of images were used to train the PLS channels themselves as well as the 3D ml-CHO. We have shown in Section 5.3.2 that the highest task SNR was achieved with the combination of $N_s = 5$ and $N_c = 15$ (Figure 5.4), but we admit that the limitations of the model observer (e.g., poorer quality of 3D PLS channels) may have contributed to the decreased performance at larger values of both N_c and N_s. One alternative approach that may help improve the quality of PLS channels is to handle each of the N_s slices separately (i.e., estimate individual sets of 2D PLS channels for each slice), and then generate the corresponding channelized image of the slice. The covariance and mean of the concatenated channelized images of the N_s slices can be used to compute decision variables by incorporating the correlation between the individual slices through the model observer mechanism rather than through efficient 3D channels which are not yet identified.

We also recognize the following limitation of the simulation study. Only four local VOIs, rather than the whole breast phantom, were extracted to represent the 3D DBT imaging data that was used in the model observer calculation. Factors such as the actual distances between lesions were not taken into account in the decision variable estimation. This shortcoming may be critical for MC cases in which lesions were more likely to be far apart from each other. It may have made the task of detecting MC lesions easier than it actually is, especially when the breast lesions were located in local anatomical backgrounds that help them stand out (e.g., less dense regions, more uniform distributed breast tissues). Moreover, when extracting the four local VOIs, we always made each of the breast lesions located in the center of each local VOI (i.e., the locations of the lesions (if any) were exactly known), However, in clinical practice, this is rarely the case because lesion locations could vary across patients, and lesions could be of significantly different characteristics (e.g., shape, size, symmetry). Hence, it would be helpful for the model observer to include additional mechanisms for handling such characteristics as lesion shape and location uncertainty (e.g., [7, 105, 116]).

All the images in the simulation study were scanned with the same DBT system geometry, and then reconstructed using the same algorithm and settings. There exist many other DBT system geometries, such as Hologic Selenia DBT system (Hologic, Bedford, MA) that uses a 15-degree angular span and 15 projection views per scan, and Siemens MAMMOMAT system (Siemens Healthineers, Erlangen, Germany) that uses a 50-degree angular span, and 25 projection views per scan. These different DBT system designs can result in different image characteristics such as spatial resolution (e.g., [166, 66]) and contrast-to-noise ratios (e.g., [84]). Similarly, various reconstruction algorithms, such as the simultaneous algebraic reconstruction technique (e.g., [84, 164]), and the total variation minimization method (e.g., [134, 48]), may produce DBT images of various properties. Hence, more comprehensive evaluations that cover the key factors in DBT acquisition, processing and presentation would be desirable to reach more definitive conclusions on the practicability and robustness of the 3D ml-CHO model observer.

5.5 Summary

In this chapter, we present a novel 3D model observer that allows for the evaluation of DBT in the detection of MF and MC breast cancers. The

3D ml-CHO model observer was used to detect multiple lesions in simulated 3D DBT images of digital breast phantoms with multiple embedded synthetic breast lesions. We have shown that the ml-CHO using 3D PLS channels could achieve high detection performance with a small number of channels, and the PLS channels significantly outperform the counterpart 3D LG channels. We have also shown that incorporating locally varying anatomical backgrounds and their correlations is helpful for making multi-lesion detection decisions. We have also discussed that the optimal design of the model observer should be adjusted when the task of clinical interest changes, or when the number of training data are limited. Together, these results demonstrate that the 3D ml-CHO could be a useful tool in task-based assessment of DBT image quality for cases with multiple lesions. In the next chapter, we employ the 3D ml-CHO to identify DBT system geometries that are most effective for the detection of MFMC. We investigate whether the consideration of multiple breast lesions may present distinct challenges to DBT system optimization.

Chapter 6

Optimizing DBT for Detecting MFMC

6.1 Introduction

Chapter 4 highlights the importance of including MFMC detection and diagnosis in the optimization of DBT. In Chapter 5, we present a practical 3D volumetric ml-CHO model observer for detecting multiple lesions in volumetric image data. In this chapter, the proposed 3D ml-CHO is employed to identify DBT system geometries that are most effective for the detection of MFMC [1,2]. The task of the 3D ml-CHO was to detect multiple synthetic lesions on 3D DBT images, and its detection performance was used to rank multiple DBT system geometries. As the design choices of narrow- or wide-arc geometry, and number of projections per scan are of particularly interest in image acquisition [127], we simulated DBT scanners of four different geometries. Digital

---

[1] **Gezheng Wen**, Subok Park and Mia K. Markey, Digital breast tomosynthesis for detecting multifocal and multicentric breast cancer: influence of acquisition geometry on model observer performance in breast phantom images. *SPIE Medical Imaging*, **10136**: 101360V, 2017. (Gezheng Wen is the primary author who developed the methods, performed the analysis, and prepared the manuscript. Mia K. Markey and Subok Park helped with the study design and manuscript revisions.)

[2] **Gezheng Wen**, Mia K. Markey, Tamara Miner Haygood, and Subok Park. Digital breast tomosynthesis for multi-lesion detection in the presence of anatomical noise. *in preparation* (Gezheng Wen is the primary author who developed the methods, performed the analysis, and prepared the manuscript. Mia K. Markey, Tamara Miner Haygood and Subok Park helped with the study design and manuscript revisions.)

breast phantoms were scanned by each of these simulated scanners. Multiple synthetic lesions were embedded into different breast regions to simulate MF and MC cases. In the preliminary version of this study presented and published in the proceedings of *SPIE Medical Imaging 2017* [153], we presented the preliminary results with multiple, identical spheres as simulated lesions. We showed that the rank order for detecting MFMC may depend on the task of clinical interest (i.e., image-level or location-level). In the extended journal manuscript in preparation for a peer-reviewed journal [148], we further explored these questions with realistically shaped breast lesions. The evaluation results highlight the performance trends of the 3D ml-CHO with different DBT geometries. We compared the rank order of the DBT geometries for MF cases to that for MC cases. We compared the rank order for detecting MFMC to the rank ordering when only unifocal cases were considered. In these two works, Gezheng Wen developed the methods, performed the analysis, and prepared the manuscript. Mia K. Markey, Tamara Miner Haygood, and Subok Park helped with the study designs and manuscript revisions.

This chapter is organized as follows: Section 6.2.1 describes the simulation of MFMC image dataset, including information on the simulated DBT scanners of different system geometries in Section 6.2.1.1, structured image background in Section 6.2.1.2, and synthetic breast lesions in Section 6.2.1.3. In Section 6.2.2, we briefly describe the employed 3D ml-CHO with PLS channels, and in Section 6.2.3, the strategies for comparing DBT system geometries.

6.2 Materials and Methods

6.2.1 Image Dataset

6.2.1.1 Simulated DBT scanner

In this study, we employed the same simulated DBT system as described in Chapter 5.2.1.1, which features an ideal photon-counting flat panel detector and a point x-ray source. The x-ray scatter in photon transport was ignored, and only quantum noise was considered. 20 keV monochromatic x-ray energy with a fixed total exposure in all scan settings was simulated. Reconstruction slice thickness was 1 mm and the in-plane resolution was 125 μm. To avoid the need to change too many different parameters from one testing condition to another, we used the conventional FBP with a Hann filter. In the case of single-lesion detection under the assumption of stationary background statistics, the FBP with a Hann filter should be appropriate for DBT optimization [160]. However, the impact of different reconstruction algorithms on the detection of MFMC cancers still remains to be seen as it is critical to consider the non-stationary nature of data statistics in such cases. We briefly explored other alternative reconstruction algorithms, including the maximum-likelihood method (e.g., [119]), the simultaneous algebraic reconstruction technique (e.g., [84, 164]) and the total variation minimization method (e.g., [134, 48]). The same evaluation procedures as discussed in the following section 6.2.3 were used. The preliminary results, as the examples shown in Appendix B.1, demonstrate that the FBP-based comparisons of the geometries may generalize to DBT systems that employ different reconstruc-

Table 6.1: List of four DBT system geometries compared in the study

Angular span (degrees)	**Number of projections per scan**	**Angular increment (degrees)**	**DBT systems with similar geometries**
24	9	3	GE SenoClair system, $(AS25, Np9)$
24	13	2	Planmed Clarity3D, $(AS30, Np15)$
60	21	3	GE2 research unit, $(AS60, Np21)$
60	31	2	Siemens MAMMOMAT, $(AS50, Np25)$

tion algorithms.

As a starting point to optimize DBT for the detection of MFMC, we investigated two key factors in the design of the acquisition process [127]: 1). angular span, and 2) number of projections per scan. As given in Table 6.1, we evaluated four DBT system geometries, two of which use a wide-arc geometry (i.e., 60 degree of angular span) while the other two use a narrow-arc geometry (i.e., 24 degree of angular span). Similarly, two of them use the uniform angular increment of 2 degrees while the other two use the uniform angular increment of 3 degrees. The geometries were represented in the format of (degree of angular span, number of projections per scan) e.g., $(AS24, Np9)$. As shown in Figure 6.1, the acquisition is symmetric with respect to the central projection angle, and the angular increment between two consecutive projection views is constant. These four system geometries cover a representative range of the

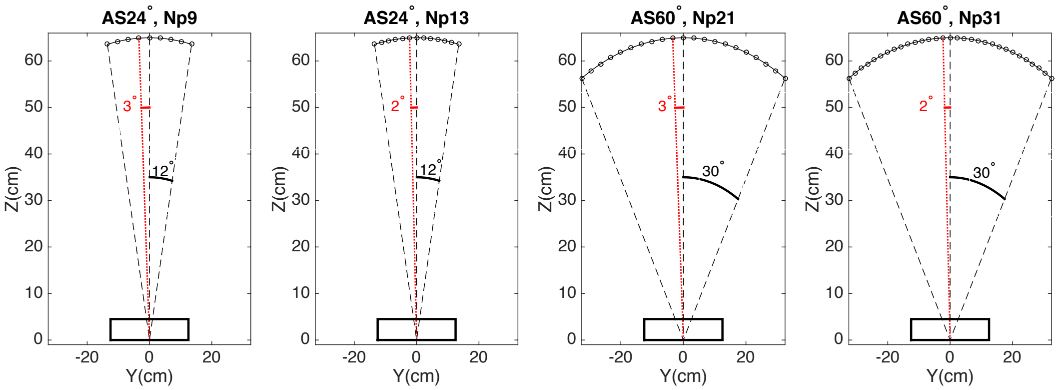


Figure 6.1: Schematic of DBT system geometries. The title of each subfigure indicates the system geometry in the format of (angular span, number of projections per scan). The angular interval between two consecutive projections (e.g., *red dotted* line and *black dashed* line) indicates the angular increment of the geometry.

geometries of current clinical, prototype, and research DBT systems (e.g., [127, 66, 57, 130]) (Table 6.1). For example, the geometry $(AS24, Np9)$ is similar to the GE SenoClair DBT system (General Electric Healthcare, Milwaukee, WI) that uses 25-degree angular span and 9 projection views [127].

6.2.1.2 Structured image background

As introduced in Chapter 5.2.1.2, digital breast phantoms mimic breast anatomy, and random structural variations between patients. We adapted the ensemble of 5000 lesion-free phantoms described in Chapter 5.2.1.2 [153, 149]. The breast phantom of size $204.5 \times 65 \times 51.5$ mm^3 with the isotropic voxel size 0.125 mm was set with a volume of 450 ml (i.e., cup size B), compressed thickness of 51.5 mm, and volumetric glandular fraction of 25%. The volumes of random 3D power-law noise were randomly added to certain local breast regions to simulate small-scale anatomical structure [153]. Figure 6.2 shows example reconstructed slices of the same breast phantom when scanned with the four DBT system geometries as described in Section 6.2.1.1.

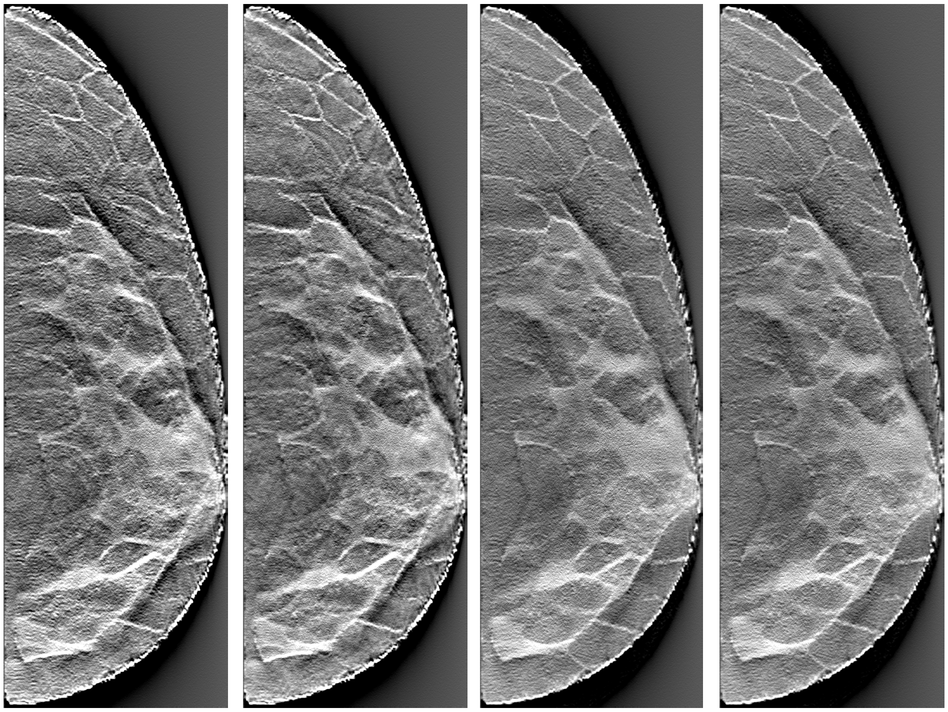

(a) AS24,Np9 (b) AS24,Np13 (c) AS60,Np21 (d) AS60,Np31

Figure 6.2: Example of reconstructed image slices of the same lesion-free breast phantom as scanned by the four simulated DBT system geometries being investigated. The same total radiation exposure was used across the geometries, and the images were displayed with the same window settings. (a) and (b) with the narrow-arc geometries shows better in-plane resolution than (c) and (d) with the wide-arc geometries, whereas (c) and (d) exhibit less tissue overlap than (a) and (b).

6.2.1.3 Synthetic breast lesions

We adapted the set of simulated synthetic 3D breast lesions as described in Chapter 5.2.1.3. The voxel size of the lesions was set at 125 μm (i.e., the same voxel size of the breast phantom). The central volumes of the well-circumscribed lesions had an approximate diameter of 6 mm ± 1 mm, and an approximate volume of 80 mm^3 ± 30 mm^3. We also embedded random spicules into the simulated spiculated lesions. The four types of lesions used in the evaluation study are shown in Figure 5.1.

The same approach as described in Chapter 5.2.1.3 was used to embed a random number of breast lesions in the phantoms for simulated MF and MC cases. The lesion locations were sampled from a prevalence distribution $\{\frac{1}{12}, \frac{2}{12}, \frac{4}{12}, \frac{5}{12}\}$ (Figure 5.2(d)). The number of lesions was sampled from a truncated Poisson distribution with a mean of 1.

(i). For MF cases, adjacent lesions were separated from each other by 30 mm (Figure 5.2(b)). Four local VOIs of in-plane size 20 $mm \times 20$ mm at each lesion location were extracted, and then combined into a large 2×2 VOI, denoted as $\mathrm{VOI}_{\mathrm{MF}}$ (Figure 5.2(e)).

(ii). For MC cases, lesions were located across different breast quadrants (Figure 5.2(c)). Similarly, four local VOIs at each lesion location were extracted, and combined into a large 2×2 VOI, denoted as $\mathrm{VOI}_{\mathrm{MC}}$ (Figure 5.2(g)).

6.2.2 3D Multi-lesion CHO with 3D PLS channels

In Chapter 5, we present a 3D volumetric ml-CHO model observer that is able to accurately detect multiple breast lesions on DBT reconstructed image slices. In this study, we employed the 3D ml-CHO with 3D PLS channels for assessing different DBT geometries [153, 148]. Following the model described in Chapter 5.2.2, there were three main steps for the 3D ml-CHO to generate the detection variables: 1) concatenate the stack of relevant reconstructed image slices together to represent the 3D DBT image data $\mathbf{X}$; 2) compute the location-level 3D PLS channels $\mathbf{T}_{\mathrm{loc}}$ [151, 152]; and 3) use the estimated PLS $\mathbf{T}_{\mathrm{loc}}$ channels in the 3D ml-CHO for generating decision variables [114].

We estimated 3D PLS channels directly from training images (i.e., $\mathbf{T}_{\mathrm{loc}} = \mathbf{PLS}(\mathbf{X}, \mathbf{Y}^{\mathrm{L}})$, where $\mathbf{PLS}$ is the PLS channel generating function (Chapter 3.2.3)that maximizes the covariance between $\mathbf{X}$ and the location-specific truth labels $\mathbf{Y}^{\mathrm{L}}$ [152]). By design, the PLS channels were naturally adaptive to complex anatomical backgrounds and signal statistics. Figure 6.3 shows example 3D $\mathbf{T}_{\mathrm{loc}}$ PLS channels estimated for MF cases with the DBT geometry $(AS24, Np13)$ (Figure 6.3(a)), and for MC cases with the geometry $(AS60, Np31)$ (Figure 6.3(b)). In this example, each of the VOI contained 5 slices (i.e., $N_s = 5$), but only the first 15 channels (i.e., the number of channels $N_c = 15$) for the central slice of the VOI are plotted. For both MF and MC cases, the following three observations could be made. First, $\mathbf{T}_{\mathrm{loc}}$ captured the characteristics of both non-spiculated and spiculated lesions. Second, lesion contrasts and lesion-presence patterns (i.e., which of the lesions were more likely to be present together in the same case) in the $\mathbf{T}_{\text{loc}}$ channels reflected lesion prevalence at different lesion locations, and the interactions between locations. Third, $\mathbf{T}_{\text{loc}}$ were adaptive to local background statistics. Hence, $\mathbf{T}_{\text{loc}}$ were useful as the common set of channels for the 3D ml-CHO to generate both image-level and location-specific decisions [152].

The linear discriminant template $\mathbf{w}_{\text{loc}}^{\text{L}}$ (as a function of $\mathbf{T}_{\text{loc}}$, $\mathbf{X}$, and $\mathbf{Y}^{\text{L}}$, were trained for computing decision variables $\mathbf{S}_{\text{loc}}^{\text{L}}$: $\mathbf{w}_{\text{loc}}^{\text{L}} = \mathbf{K}_{\mathbf{V}}^{-1}\bar{\mathbf{V}}$, $\mathbf{S}_{\text{loc}}^{\text{L}} = (\mathbf{w}_{\text{loc}}^{\text{L}})^{T}(\mathbf{T}_{\text{loc}}^{T}\mathbf{X})$, where $\mathbf{V} = \mathbf{T}_{\text{loc}}^{T}\mathbf{X}$ (^{T} indicates transpose) is the set of channelized images that are classified into $\mathbf{V}_1$ and $\mathbf{V}_0$ based on $\mathbf{Y}^{\text{L}}$ (i.e., $\mathbf{V}_i(j)$ are the columns that $\mathbf{Y}^{\text{L}}(j)$ is i, where $i = 0$ or 1, $j = 1, 2, .., N_l$), $\bar{\mathbf{V}} = \bar{\mathbf{V}}_1 - \bar{\mathbf{V}}_0$ is the mean channelized signal(s), and $\mathbf{K}_{\mathbf{V}} = \frac{1}{2}(\mathbf{K}_{\mathbf{V}_1} + \mathbf{K}_{\mathbf{V}_0})$ is the mean covariance matrix of $\mathbf{V}_1$ and $\mathbf{V}_0$ [61]. The higher the score is, the more likely the lesion is present. We note that $\mathbf{S}_{\text{loc}}^{\text{L}}$ contained location-specific scores that were used for making both image-level and location-specific decisions. For image-level decisions, a second-stage linear decision template was used to convert $\mathbf{S}_{\text{loc}}^{\text{L}}$ into scalar image-level scores [152, 114].

6.2.3 Evaluation

Adapting the evaluation strategies in Chapter 5.2.3 for 3D ml-CHO, we used two FOMs to measure observer performance: 1) a task SNR for image-level decision variables [61]. 2). $\text{AUC}_{\text{AFROC}}$ over $[0, 1]$ for location-specific de-

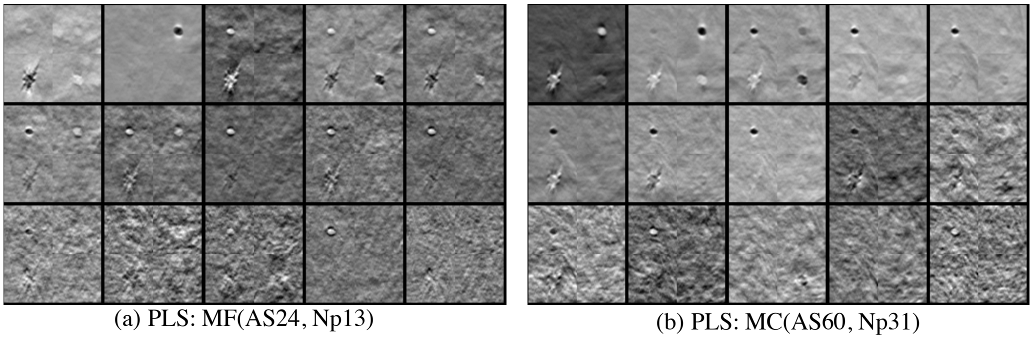

(a) PLS: MF(AS24, Np13) (b) PLS: MC(AS60, Np31)

Figure 6.3: The first 15 columns of example 3D PLS $\mathbf{T}_{\text{loc}}$ channels: For each subfigure, 1st row shows the 1st (*left*) to 5th (*right*) channel, 2nd row shows the 6th (*left*) to 10th (*right*) channel and etc. (a) $\mathbf{T}_{\text{loc}}$ for MF cases with the geometry $(AS24, Np13)$; (b) $\mathbf{T}_{\text{loc}}$ for MC cases with the geometry $(AS60, Np31)$. PLS $\mathbf{T}_{\text{loc}}$ in (a) and (b) captured lesion characteristics. Lesion contrasts and lesion-presence patterns in $\mathbf{T}_{\text{loc}}$ also reflected lesion prevalence at different locations, and the interactions between locations.

cision. For each run of the evaluation experiment, we first embedded a random number of lesions to be detected, ranging from 0 to 4, into the 5000 lesion-free structured breast phantoms. Then stratified random sampling [38] was used to randomly divided these simulated MFMC cases into three non-overlapping phantom groups: 2000 cases for estimating the 3D PLS channels $\mathbf{T}_{\text{loc}}$, 2000 for training the linear discriminant template $\mathbf{w}_{\text{loc}}^{\text{L}}$, and the other 1000 cases for testing. To investigate the impact of N_s on MFMC cancer detection, the 3D ml-CHO model was trained independently with six different sets of $\mathbf{X}$ (i.e., $\mathbf{X}$ with $N_s = 1, 3, 5, 7, 11$, or 15), resulting in six different 3D ml-CHOs. The experiment was repeated for 20 runs, where each run used a different division of the three groups. The average and two standard deviations of the task SNR and $\text{AUC}_{\text{AFROC}}$ were reported to identify performance trends.

We compared the rank orders for the detection of MFMC lesions to those of unifocal lesions. The same ensemble of lesion-free phantoms as described in Section 6.2.1.2 were used as the image background. We simulated two sets of unifocal cases by choosing the local VOI 2 and VOI 4 of the MF cases as the potential lesion locations, respectively (Figure 5.2(e)). The same set of lesions as described in Section 6.2.1.3 was used, but each lesion-present case only contained one lesion. To detect the unifocal lesion, a typical 3D one-lesion CHO with 3D PLS channels [81, 114, 107] was used. This could be considered as a special case of the 3D ml-CHO with the number of possible lesion location $N_l = 1$ [152, 149]: 1) the channels $\mathbf{T}$ were estimated with the binary truth label $\mathbf{Y}^{\mathrm{I}}$, 2) the decision template $\mathbf{w}$ was of size $N_c \times 1$, and 3) the decision variable S for each image was a scalar.

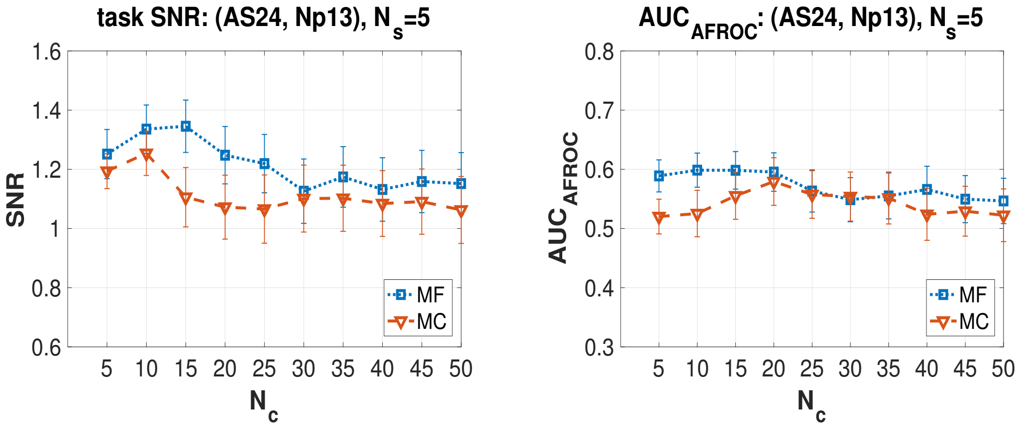
6.3 Results

6.3.1 FOMs for MF versus MC

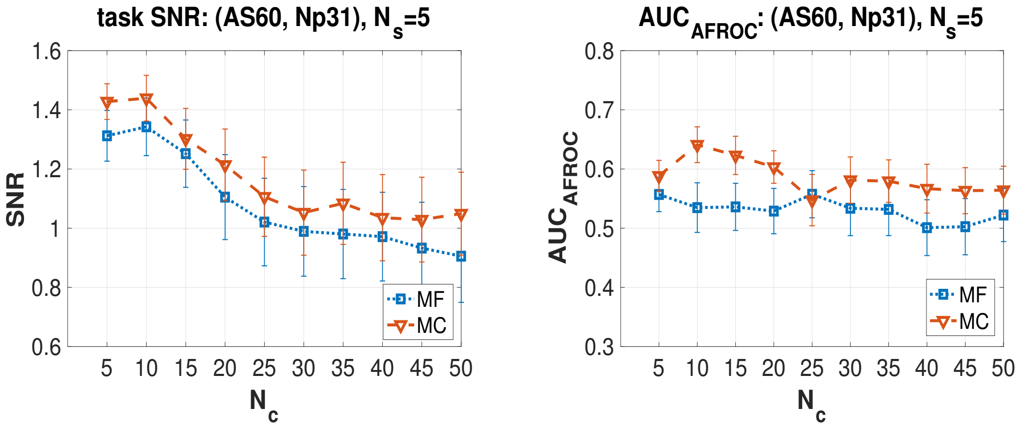
Figure 6.4 shows comparisons of the 3D ml-CHO performance between the MF and MC cases. The horizontal axis of the plots is the number of PLS channels N_c used in the model observer. The number of reconstructed slices N_s in the 3D image data $\mathbf{X}$ was set to 5. For the geometry $(AS24, Np13)$, both the task SNR (Figure 6.4(a)) and $\mathrm{AUC}_{\mathrm{AFROC}}$ (Figure 6.4(b)) for the MF cases (*blue*) were higher than those for the MC cases (*red*). In contrast, for the geometry $(AS60, Np31)$, both the task SNR (Figure 6.4(c)) and $\mathrm{AUC}_{\mathrm{AFROC}}$

(Figure 6.4(d)) for the MF cases (*blue*) were lower than those for the MC cases (*red*). These results suggest that the narrow-arc geometry may be more effective for detecting MF lesions while the wide-arc geometry may be more effective for MC lesions.

The aforementioned difference in the performance trends could be partially explained by the fact that the local VOIs in the MC cases were located in different breast quadrants, and hence local background statistics of the local VOIs were more dissimilar in the MC cases than in the MF cases. Many studies found that the power law exponent β of anatomical NPS was positively correlated with breast density [34, 85, 90]. In order to describe the variations of local background textures, we estimated β values of the lesion-free VOIs from the 5000 phantom ensemble. As is typical in similar studies (e.g., [79, 34, 46]), only the central slice of the local VOIs was used to compute the 2D NPS, and β was estimated by fitting a power-law to the radially averaged 1D NPS. The average and standard deviations of β from 2000 local VOIs were estimated for all four DBT system geometries. The example results with the geometry $(AS24, Np13)$ in Table 6.2 shows that the variation in the estimated β values for the MF cases was smaller than that of the MC cases. For the four local VOIs in MF cases, all β values were approximately 2.4, while the local VOIs in MC cases had significantly different β values. For example, the estimated value of β for VOI 4 was 2.83, while the estimate value of β for VOI 1 was only 2.12. In addition, we measured spatial correlations between the voxels with



(a) task SNR $(AS24, Np13)$ (b) AUC$_{\text{AFROC}}$ $(AS24, Np13)$



(c) task SNR $(AS60, Np31)$ (d) AUC$_{\text{AFROC}}$ $(AS60, Np31)$

Figure 6.4: FOM comparisons between MF cases and MC cases. 5 reconstructed slices were used by the 3D ml-CHO. (a)-(b): The DBT geometry $(AS24, Np13)$ may be more effective for detecting lesions in MF cases (blue) than MC cases (red). (c)-(d): The DBT geometry $(AS60, Np31)$ may be more effective for MC cases (red) than MF cases $(blue)$.

Table 6.2: The mean and standard deviations of β of NPS of local VOIs in MF and MC cases using the images with the geometry $(AS24, Np13)$.

	VOI 1	VOI 2	VOI 3	VOI 4
MF cases	2.43 ± 0.13	2.32 ± 0.10	2.47 ± 0.16	2.42 ± 0.12
MC cases	2.12 ± 0.09	2.41 ± 0.15	2.63 ± 0.11	2.83 ± 0.23

the four VOIs of MF cases, and of MC cases. Figure 6.5 plots the example correlation estimated with images of lesion-free phantoms acquired with the DBT geometry $(AS24, Np13)$. It can be observed that for MC cases shown in Figure 6.5(b), the spatial correlations within each of the local VOIs were noticeably different. Among the four local VOIs, local VOI 1 had the shortest correlation length, but the largest variation in intensity along the dominant direction of intensity changing, while local VOI 4 had the most smoothly-distributed intensity, but the longest and strongest correlations. Local VOI 2 and 3 were in between, closer to local VOI 1 and 4, respectively. This suggests that as the local VOIs were located in different breast regions, and well separated, local anatomical noise in MC cases may be significantly different across the local VOIs, affecting the difficulty of MC lesion detection tasks. In contrast, for MF cases shown in Figure 6.5(a), the spatial correlations of the local VOIs of MF cases were similar in terms of both the correlation intensity and length, which suggests that spatial distribution of breast tissue were similar in these local breast regions. Similar trends are observed with the other three geometries.

When comparing Figure 6.4(a) with Figure 6.4(c), and Figure 6.4(b) with Figure 6.4(d), when $N_c < 15$, both MC and MF cases were better de-

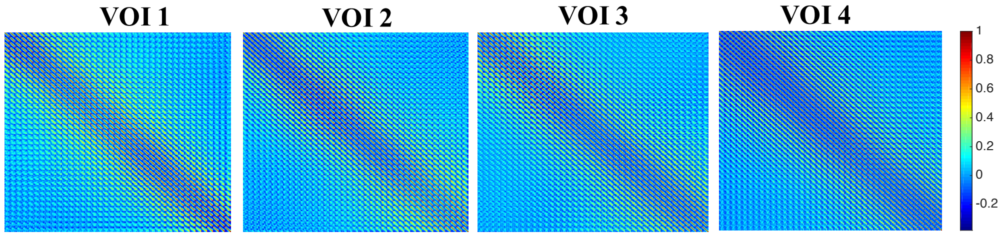
Table 6.3: Task SNR difference between $(AS60, Np31)$ and $(AS24, Np13)$ (i.e., $SNR(AS60) - SNR(AS24)$). When $N_c < 15$, for both MF and MC cases, SNR of $AS60$ was higher with than that of $AS24$.

	$N_c = 5$	$N_c = 10$	$N_c = 15$	$N_c = 25$	$N_c = 35$	$N_c = 50$
MF cases	0.06	0.10	-0.08	-0.19	-0.15	-0.14
MC cases	0.23	0.18	0.19	0.04	-0.02	-0.02

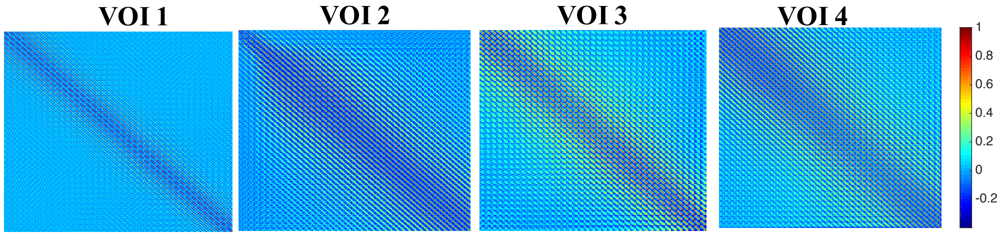
tected with the wide-arc geometry $(AS60, Np31)$ than the narrow-arc geometry $(AS24, Np13)$. Table 6.3 summaries the difference in task SNR from $(AS24, Np13)$ to $(AS60, Np31)$ with respect to N_c. When $N_c < 15$, the SNR differences were positive for both rows (i.e., a higher SNR for $(AS60, Np31)$ than $(AS24, Np13)$). Comparing the magnitude of SNR differences for MF cases (*top row*) and MC cases (*bottom row*), the wide-arc geometry $(AS60, Np31)$ improved the detection of MC cases much more than MF cases (e.g., for $N_c = 5$, SNR increased by 0.23 for MC cases against 0.06 for MF cases). This may have contributed to the reversal of detectability ranking between the two cases. However, as shown in Table 6.3, for $N_c \geq 15$, this was no longer true because the uncertainty, introduced by factors such as the dimension of covariance (e.g., 40×40 for $N_s = 5, Nc = 40$) and the limited sample size (e.g., 2000), likely led to reduce the absolute performance levels of the 3D ml-CHO while still maintaining performance trends.

6.3.2 FOMs for Different DBT Geometries

Figure 6.6(a)-(b) plot the FOM comparisons across the four DBT geometries for MF cases. 15 PLS channels (i.e., $N_c = 15$) were used by each



(a) MF cases

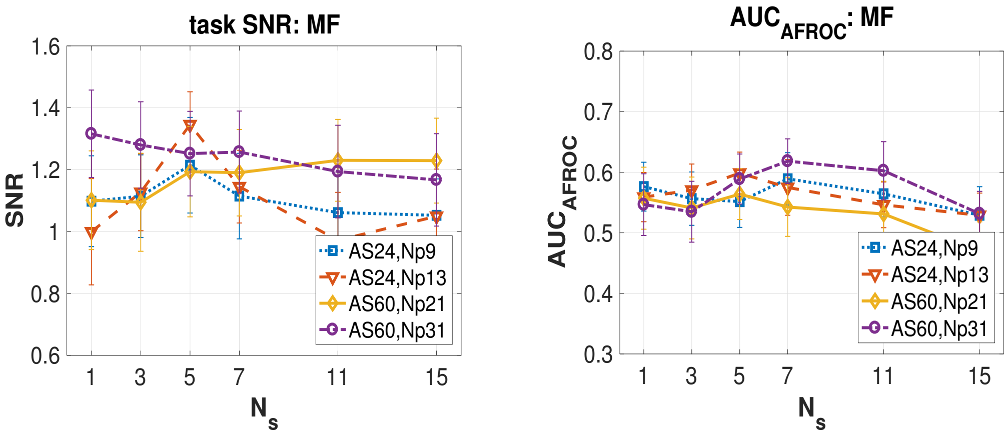


(b) MC cases

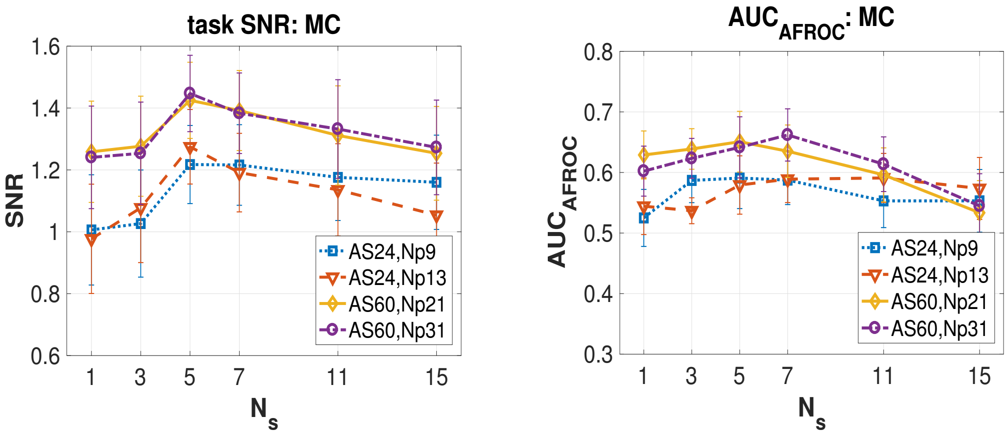
Figure 6.5: (a)-(b) Spatial correlations between the pixels within the four local VOIs of MF cases and of MC cases. The calculations were done on the reconstructed slices of lesion-free phantoms acquired with the DBT geometry $(AS24, Np13)$. For the MC cases, the spatial correlations were noticeably different across the local VOIs, while for MF cases, they were similar.

of the differently trained four model observers. Figure 6.6(c)-(d) plot the counterpart results for MC cases. For both wide- and narrow arc-geometries ($AS60$ and $AS24$), changing the number of projections Np changed the trends of detection performance for MF cases. For example, in Figure 6.6(a) of task SNR, the geometry $(AS24, Np13)$ (red) has a higher SNR than the geometry $(AS24, Np9)$ $(blue)$ when $N_s = 5$, but a lower SNR when $N_s = 11$. This difference could be explained by the fact that for the same angular span and the same total amount of radiation exposure, increasing Np improves the angular sampling, but also results in higher quantum noise due to the reduced dose for each of the projections. Hence, with a relatively large N_s, the detection performance with the geometry $(AS24, Np13)$ may have been significantly degraded due to the overwhelming noise. However, this is not observed for MC cases: changing Np did not noticeably change the trends of detection performance (Figure 6.6(c)-(d)). For example, in Figure 6.6(c) of task SNR, the geometries $(AS60, Np21)$ (*yellow*) and $(AS60, Np31)$ (*purple*) were approximately overlapped. This could be explained as discussed in Section 6.3.1: locally varying anatomical backgrounds in MC cases, rather than angular sampling or quantum noise, have the dominating impact on the detection of MC lesions.

Moreover, the rank orders of the four geometries by the task SNR or $\text{AUC}_{\text{AFROC}}$ may not be the same. This is observed with both MF and MC cases. For example, for $N_s \geq 7$, the geometry $(AS60, Np21)$ (*yellow*) was significantly better than the geometry $(AS24, Np13)$ (*blue*) by task SNR (Figure 6.6(a)), but worse by $\text{AUC}_{\text{AFROC}}$ (Figure 6.6(b)). This was also closely related to the discussions in Section 6.3.1, where we have shown that the wide-arc geometries may better help improve the detection of MC cases. Although the geometry $(AS60, Np21)$ may be better for making detection decision at the coarse image-level, it may lack the ability for accurate location-specific decisions due to potentially affecting factors such as relatively higher anatomical noise, lower in-plane resolution and etc. Therefore, the optimal DBT geometry for MFMC detection may vary as the clinical task of interest changes (e.g., find at least lesion versus find the exact number of lesions).



(a) task SNR for MF cases (b) $\text{AUC}_{\text{AFROC}}$ for MF cases



(c) task SNR for MC cases (d) $\text{AUC}_{\text{AFROC}}$ for MC cases

Figure 6.6: (a)-(b): FOM comparisons in MF cases across the four DBT geometries. 15 PLS channels (i.e., $N_c = 15$) were used by each of the differently trained four model observers. (c)-(d): the counterpart FOM comparisons in MC cases. For both wide- and narrow arc-geometries ($AS60$ and $AS24$), changing Np changed the trends of detection performance in MF cases, but not in MC cases. The rank orders of the four geometries may not be the same by the task SNR or $\text{AUC}_{\text{AFROC}}$ in MF cases or MC cases.

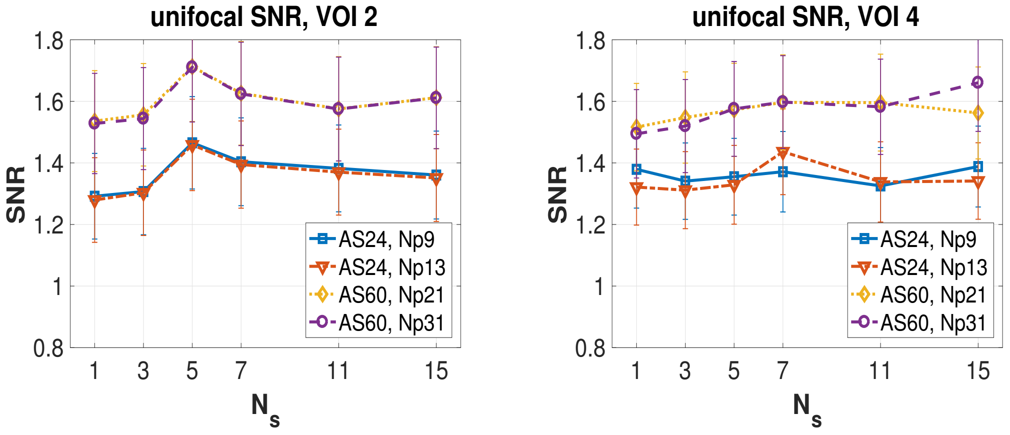
6.3.3 FOMs for MFMC versus Unifocal

Figure 6.7 shows the rank order of the four geometries when only unifocal cases were considered (i.e., at most one lesion was present). 15 3D PLS channels was used in each of the differently trained one-lesion CHOs. From the results shown in Figure 6.7, the following observations could be made:

(i). For each of the two chosen local VOIs, the unifocal SNR of $AS60$ (*yellow* and *purple*) was higher than that of $AS24$ (*blue* and *red*). Hence, for these particular settings, the wide-arc geometry may be consistently better than the narrow-arc geometry, which was different from the rank order of task SNR for MF cases (Figure 6.6(a)).

(ii). For each of the two chosen local VOIs, changing Np for a fixed angular span did not influence the trends of detection performance. For example, unifocal SNR for $(AS60, Np21)$ (*yellow*) and for $(AS60, Np31)$ (*purple*) were overlapped in Figure 6.7(a). This differs from the trends of task SNR for MF cases (Figure 6.6(a)).

(iii). The trend of unifocal SNR as a function of N_s was different for the two local VOIs. In Figure 6.7(a) for VOI 2 of MF cases, all of the four geometries achieved the highest SNR when N_s was 5, and increasing or decreasing N_s led to a lower SNR. This was generally consistent with the trend of task SNR for MF cases (Figure 6.6(a)). However, in Figure 6.7(b) for VOI 4 of MF cases, when N_s increased, the unifocal SNR



(a) unifocal SNR

(b) unifocal SNR

Figure 6.7: (a)-(b): FOM comparisons in MF cases across the four DBT geometries. 15 PLS channels (i.e., $N_c = 15$) were used by each of the differently trained four model observers. (c)-(d): the counterpart FOM comparisons in MC cases. For both wide- and narrow arc-geometries ($AS60$ and $AS24$), changing Np changed the trends of detection performance in MF cases, but not in MC cases. The rank orders of the four geometries may not be the same by the task SNR or $\mathrm{AUC}_{\mathrm{AFROC}}$ in MF cases or MC cases.

generally increased for $AS60$ (*yellow* and *purple*), but the SNR peaked at $N_s = 7$ for $AS24$ (*blue* and *red*).

Thus, DBT geometries informative for unifocal cases may not be informative for detecting MFMC breast cancer. Similarly, similar results and trends as for MF cases are observed with MC cases: the rank orders for MC cases and for unifocal cases may be different in terms of task SNR.

6.4 Discussion

In this study, we used the 3D volumetric ml-CHO model with 3D PLS channels to detect multiple lesions in simulated DBT reconstructed image slices. We investigated two key factors in the design of DBT image acquisition: angular span and the number of projections. We compared the effectiveness of four representative DBT geometries for detecting MF and MC breast lesions. It is evident from Sections 6.3.1 and 6.3.2 that MF and MC cases may raise challenging questions in DBT system design because the rank order of the four system geometries may depend on the task of clinical interest:

(i). The larger angular span of the wide-arc geometry (e.g., $AS60$) tends to decrease the in-plane spatial resolution [127], and for the same total x-ray exposure, the larger number of projections (e.g., $Np31$) resulted in higher quantum noise in the reconstructed images [130]. However, the finer angular sampling and higher depth resolution may offer benefits in handing variations in the spatial distributions of breast tissues across different local breast regions. This advantage may explain the relatively larger improvement in the detection performance for MC cases than MF cases when switching from narrow-arc geometry $AS24$ to wide-arc geometry $AS60$ (Table 6.3). This is because the presence of anatomical noise has shown to be the dominating factor that affects observer detection performance levels and trends, which is particularly significant for causing performance difference among model observers when given the same amount of training of channels and observer (e.g., [81, 152, 101, 34]). As a result of all these effects, wide-arc geometries may be more suitable for MC cases than for MF cases as the lesions in MC cases are physically separated from each other, and influenced by different levels of local anatomical noise.

(ii). On the other hand, when multiple lesions, as in MF cases, are close to each other, the appearance of one lesion may affect the detection of another lesion. For example, out-of-focus blurring, or complex structures of a spiculated lesion may superimpose with another lesion that is in focus. Higher in-plane resolution may be required to distinguish fine anatomical details inside a relatively small region of breast, hence the narrow-arc geometries may have led to better detection accuracy for MF cases than for MC cases.

In this study, the x-ray scatter in photon transport was ignored, and only quantum noise was simulated. Prior studies (e.g., [83, 156, 129, 49]) have shown that scattered x-rays significantly degrade image quality of DBT reconstructed images, including reducing the contrast, increasing noise, and reducing the signal-difference-to-noise ratio of a lesion. They may also introduce artifacts, and reduce quantitative accuracy of attenuation values. Hence, including Monte Carlo simulations [18, 132] in the image acquisition would be valuable to study whether scattering may influence lesion detection in MFMC cases. This may also help improve the development of scatter correction algorithms. Moreover, in this study, we only chose one radiation exposure level that matched with the exposure level measured with real DBT scanners in similar scan settings. We have shown that different usage of the same radiation dose (i.e., different Np for the same AS) may lead to different performance trends for lesion detection in MFMC cases (Figure 6.2). In the combo acquisition mode of 2D and 3D images, FDA-approved DBT systems use approximately double the radiation dose of what is used in conventional 2D mammography. However, the optimal radiation dose for a diagnostic DBT scans, and how to efficiently use the limited available dose for best diagnostic image quality, are still ongoing research questions (e.g., [51, 69, 140]). There is also increasing concern about reducing radiation dose when designing imaging systems. Hence, it would be valuable to investigate how increased quantum noise obtained at reduced dose level may influence lesion detection in MFMC cases. The study could be easily extended by changing the total number of x-ray photons that the simulated x-ray source emits into the breast phantom. The model observer performance at varying dose levels could be directly compared. Overall, these potential extensions would also highlight the benefits of well-designed simulation studies as they are flexible, safe, and cost-efficient for evaluating DBT imaging system configurations.

We recognize the following limitation of this simulation study. The 3D ml-CHO only worked with the four extracted local VOIs, rather than the whole breast phantom. Factors such as the actual distances between lesions were not considered in the decision variable estimation. This shortcoming may be critical for MC cases in which lesions were more likely to be separated from each other. It may have made the task of detecting MC lesions easier than it actually is, especially when the breast lesions were located in local anatomical backgrounds that help them stand out (e.g., less dense regions, more uniform distributed breast tissues). Moreover, when extracting the four local VOIs, we always made each of the breast lesions located in the center of the local VOIs (i.e., the locations of the lesions (if any) were exactly known), However, in clinical practice, this is rarely the case because lesion locations could vary across patients, and lesions could be of significantly different characteristics (e.g., shape, size, symmetry). Hence, it would be helpful for the model observer to include additional mechanism for handling such lesion and location uncertainty (e.g., [105, 116, 31]).

There exist many other system geometries, such as $(AS15, Np15)$ in the Hologic Selenia DBT system (Hologic, Bedford, MA), and other system parameters besides the number of projections and the angular span for the design of DBT image acquisition. All of the four geometries in our study used uniform angular increment and equally distributed dose. However, prior studies [57, 31] have shown that variable angular increment (i.e., projection sampling is dense at the sides but sparse in the middle; or projection sampling is dense in the middle but sparse at the sides) may improve contrast-to-noise ratio of breast masses. Even for a fixed angular span and total number of projections, changing angular increment leads to a set of DBT geometries with significantly different levels of readers' perception of contrast-detail objects in DBT images [57]. Variable dose distribution may also help, e.g., by enhancing microcalcification detectability [40]. More comprehensive comparisons are necessary to reach more definitive conclusions. Nevertheless, we emphasize that our overarching clinical goal is to optimize DBT for the detection of MFMC lesions. This study was the first to investigate how the presence of multiple tumors impacts DBT system optimization. We have learned that a given DBT system geometry may not yield images that are equally informative for detecting MF, MC, and unifocal breast lesions, i.e., it may be necessary to optimize DBT for these separate tasks. Analogous to the use of different imaging protocols for clinical CT scans of different body regions (e.g., abdomen, cardivascular, brain), our study suggests that it could be valuable for a DBT system to support different image acquisition modes for different clinical scenarios, especially in diagnostic settings.

6.5 Summary

In this chapter, we present a model observer study that investigated DBT system geometries for detecting MF and MC breast cancer. The 3D volumetric ml-CHO model observer was used to detect multiple lesions in simulated 3D DBT images of digital breast phantoms with multiple embedded synthetic breast lesions. We have shown that the consideration of multiple breast lesions may present distinct challenges to DBT system optimization.

The narrow-arc geometry may be generally more effective for detecting MF cases than MC cases. In contrast, the wide-arc geometry may be generally more effective for detecting MC cases than MF cases. For the same angular span, changing the number of projections per scan may change the detection performance of the model observer. We have also shown that the rank orders of the geometries for MF and MC cases and for unifocal cases were different in terms of task SNR. Hence, the optimal geometry of DBT may vary when the task of clinical interest changes, and a given DBT system geometry may not yield images that are equally informative for detecting MF, MC, and unifocal breast cancers. The knowledge obtained through this and future studies may help improve assessment methodology in regulatory decision making for and utilization of DBT systems in the clinic.

In the next chapter, we present an ongoing human observer study with experienced breast imaging radiologists [147], through which we will validate the clinical relevance and significance of our model observer studies (Chapter 5 and 6). It is hypothesized that the rank order of DBT systems by the human observer performance would generally agree with the performance trends shown in our model observer studies.

Chapter 7

Human Observer Study of DBT Optimization for MFMC Detection

7.1 Introduction

In this chapter, we present an ongoing human observer study[1] to validate the clinical relevance of our model observer studies reported in Chapter 5 and 6. The main motivation of the study comes from the fact that our previous findings on optimizing DBT for MFMC detection, as we present in Chapter 6, were based on the performance of the 3D ml-CHO model observer under certain imaging assumptions. As introduced in Chapter 5, the 3D ml-CHO, as a potentially efficient observer, is designed to extract all available information in DBT images for determining the presence of multiple breast lesions. However, when a human observer (e.g., a radiologist) interprets a DBT case in clinical practice, he/she may not be able to effectively utilize all the image information toward his/her decision making process [159, 73, 74, 25]. For example, normal anatomical structures in the images (e.g., fibroglandular tissue

---

[1] **Gezheng Wen**, Tamara Miner Haygood and Mia K. Markey, A human observer study of multi-lesion detection in digital breast tomosynthesis. (to appear in) *Medical Image Perception Society Meeting XVII* 2017 (Gezheng Wen is the primary author who develops the methods, performs the analysis, and will deliver the oral presentation. Mia K. Markey and Tamara Miner Haygood help with the study design and presentation preparations)

in a mammogram, ribs in a chest radiograph) may camouflage abnormalities of clinical interest, and hence the abnormalities may not well perceived by humans [73]. Hence, it is necessary to validate whether the performance trends from the model observer study are consistent with the trends from the counterpart human observer study.

This study focuses specifically on the impact of different DBT system geometries on radiologists' detection performance for MFMC. Particularly, we simulate DBT images of synthetic MFMC cases with the four aforementioned geometries (Chapter 6.2.1.1). Then we ask readers to perform a multi-lesion detection task on each case of the DBT images (i.e., find all present synthetic breast lesions). Human observer performance with each of the geometries is measured from his/her accuracy and confidence, and then compared to study whether the performance trends are consistent. The study has been approved by the Institutional Review Boards of The University of Texas MD Anderson Cancer Center and The University of Texas at Austin. In the preliminary version of this study to be presented at *Medical Image Perception Society Meeting XVII* in 2017 [147], we will present the preliminary comparison results. In the extended journal manuscript in preparation for a peer-reviewed journal [148], we further explore these questions with more comprehensive evaluations and analysis.

7.2 Materials and Methods

7.2.1 Image Dataset

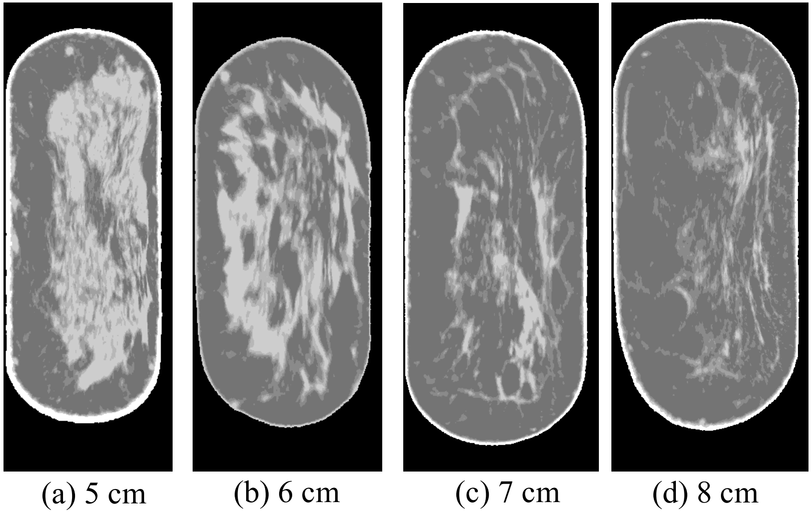
7.2.1.1 Simulated DBT scanners

In this study, we investigate the same set of simulated DBT scanners as described in Chapter 6.2.1.1, which includes four DBT geometries $(AS24, Np9)$, $(AS24, Np13)$, $(AS60, Np21)$ and $(AS60, Np31)$. Again, the x-ray scatter in photon transport is ignored, and only quantum noise is considered. $20keV$ monochromatic x-ray energy with a fixed total exposure in all scan settings is simulated. Reconstruction slice thickness is 1 mm and the in-plane resolution is 125 μm. The conventional FBP with a Hann filter is used for reconstruction.

7.2.1.2 Structured image background

For simulating the structured image background, we use a new set of patient-based, 3D anthropomorphic computational phantoms, based upon the dedicated breast computed tomography imaging data of normal patients [47]. The phantoms have been utilized in several previous studies of breast imaging system optimization and/or evaluation (e.g., [71, 68, 117, 36, 117]). With different spatial distributions of the six segmented breast tissue classes (i.e., skin, adipose, and several fractional glandular densities), the phantoms differ in volume, breast density, and local anatomical variations. The voxelized phantoms are realistically compressed using a finite element model [139] and deformed into a boundary mesh representation. 144 different breast phantoms

Figure 7.1: Coronal views of four example phantoms of (a) 5 *cm*, (b) 6 *cm*, (c) 7 *cm* and (d) 8 *cm* compressed thickness

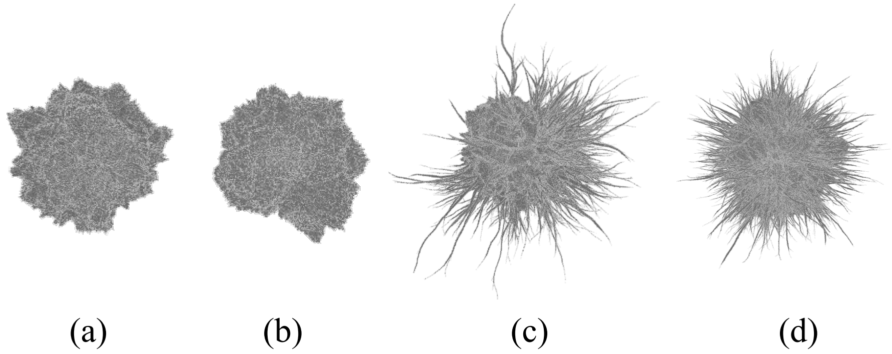

(a) 5 cm (b) 6 cm (c) 7 cm (d) 8 cm

have been generated, and they are compressed into approximately 5, 6, 7, or 8 *cm*. The phantoms are processed to have an isotropic resolution of 250 μm. Figure 7.2.1.2 shows the coronal views of four example breast phantoms of different compressed thickness.

7.2.1.3 Synthetic breast lesions

We have simulated a new set of synthetic 3D breast lesions following the framework in Ref. [135]. That is, a modified stochastic Gaussian random sphere (GRS) model is used to generate a central tumor mass, and an iterative fractal branching algorithm is used to add random complex spicule structures for simulating spiculated lesions. Specially, the GRS is used first to construct the shape of the central mass by defining a parametric surface using

Figure 7.2: The four types of simulated breast lesions used in the study (i.e., the lesions to be detected by a human observer): (a)-(b): non-spiculated lesions, (c)-(d) spiculated lesions after embedding random spiculated structures to the well-circumscribed lesions.

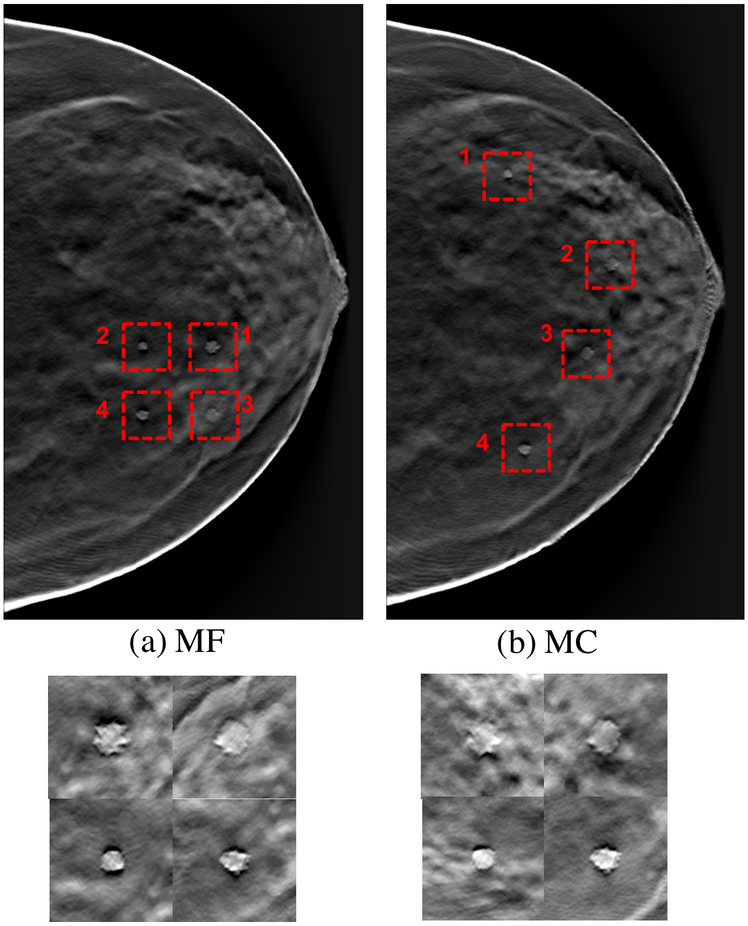


the radial distance of the surface from the origin. Varying the statistics of the GRS shapes, controlled by the maximum order of spherical harmonics and the variance of radial distance, leads to different levels of surface irregularities. Then, set of low-frequency modifications, denoted as “bumps” and “spikes”, and high-frequency modifications denoted as a “fuzzy” surface texture, are added to modify lesion surface for a greater degree of surface variations (e.g., lobulated surface, pointy surface), and hence better realism. For spiculated lesions, an iterative fractal branching algorithm recursively creates a set of segments based on a set of growing rules. Each of these segments is represented as conical frustums (i.e., characterized by a starting and an ending 3D location, a starting and an ending radius, and length). To simulate the variations across different spiculated lesions, all the growing rules include a set of random variables, such as bifurcation probability, direction of extension, emerging density and etc. In this study, we use the example combinations of user-defined parameters as listed in Ref. [135] for simulating lesions. The contrast between the lesion and the background is set to around $10 \pm 0.4\%$ to simulate the typical appearance of a lesion embedded in normal breast tissues [145]. The four types of lesions used in the evaluation study (i.e., the lesions to be detected by a radiologist) are shown in Figure 7.2.

Using the similar approaches as described in Chapter 5.2.1.3 and 6.2.1.3 for simulating synthetic MF and MC cases, lesion-present breast phantoms are generated by embedding a random number of breast lesions in the phantoms. We divide the breast into four quadrants (Figure 5.2(a)), and there are four possible lesion locations across the breast (Figure 7.3(a) for MF cases and Figure 7.3(b) for MC cases). The lesion locations are sampled from a prevalence distribution $\{\frac{1}{12}, \frac{2}{12}, \frac{4}{12}, \frac{5}{12}\}$ (Figure 5.2(d)). The number of lesions was sampled from a truncated Poisson distribution with a mean of 1.

(i). For MF cases, adjacent lesions were separated from each other by 30 mm (Figure 7.3(a)). Four local VOIs of in-plane size 20 $mm \times 20$ mm at each lesion location were extracted, and then combined into a large 2×2 VOI, designed as $\mathrm{VOI}_{\mathrm{MF}}$ (Figure 7.3(c)).

(ii). For MC cases, lesions were located across different breast quadrants (Figure 7.3(b)). Similarly, four local VOIs at each lesion location are extracted, and combined into a large 2×2 VOI, designed as $\mathrm{VOI}_{\mathrm{MC}}$ (Figure 7.3(d)).



(a) MF (b) MC

(c) $\mathrm{VOI}_{\mathrm{MF}}$ (d) $\mathrm{VOI}_{\mathrm{MC}}$

Figure 7.3: Example DBT reconstructed slices, and the corresponding whole VOI slices of two MFMC cases. The two lesion-present phantoms in (a) and (b) were scanned. (a) and (c) show an example MF case with 4 lesions that are present in the same breast quadrant. (b) and (d) show an example MC case with 4 lesions that are present in different breast quadrants. The *red dashed* boxes in (a) and (b) indicate the extracted local VOIs centered at the four possible lesion locations, where the *red digits* (i.e., 1-4) indicated the local VOI indices. All four lesions are contrast-enhanced for better visualization: the two lesions on the top (i.e., local VOI 1 and 3) are non-spiculated lesions, while the two lesions on the bottom (i.e., local VOI 2 and 4) are spiculated lesions. Lesion shapes in the local VOIs of all the simulated MFMC cases are of the same order as in (c) and (d).

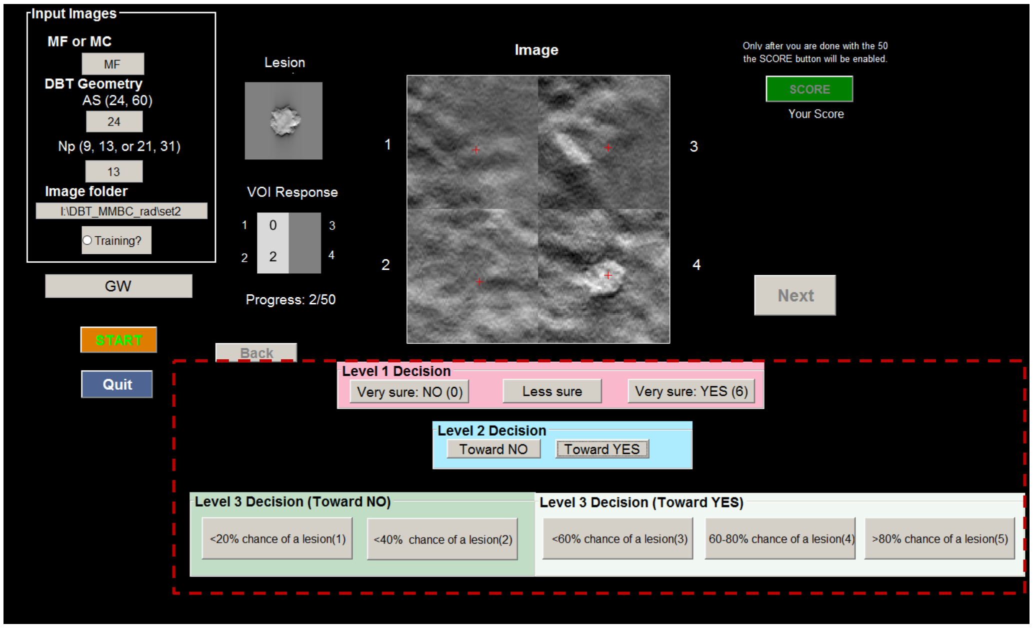
7.2.2 Experiment Design

7.2.2.1 Human observers

Five radiologists with specialty in breast imaging and experience with DBT have been recruited from The University of Texas MD Anderson Cancer Center to participate in the study. For each image, the task of the participants is to read the DBT images for detecting multiple lesions, and to determine the presence or absence of a lesion using a 0-6 ordinal scale (i.e., a score of 0 corresponds to the lowest and 6 the highest confidence of a present lesion). A seven-point scale is a good balance between having enough points of discrimination without having to maintain too many response options [98], and thus has been widely used in medical imaging perception studies to quantify observers diagnostic accuracy (e.g., [159, 111, 112]). To reduce inter-participant variation, an instruction sheet with verbal descriptions of the recommended score levels is provided. At the beginning of the study, the readers have training sessions (i.e., read 15 images not selected for the reader study) to familiarize themselves with the interface, the detection task, and the rating scale. Instant feedback is provided to the readers in the training sessions, and their performance is monitored to guarantee that at the end of the training the readers can generally follow the recommended scoring rules and achieve reasonable accuracy in the detection task.

7.2.2.2 Experimental interface

Figure 7.4 presents the schematic of the experimental interface of the study. For each DBT image set, only the central slice of the VOI is shown to the readers (i.e., $N_s = 1$). The 2D VOI is divided into four equal-size local VOIs, and each local VOI may contain at most one lesion at the center. For each local VOI, the participants need click the mouse within the local VOI and report a confidence score in the 0-6 scale for the presence of a lesion in the local VOI. A three-stage score-reporting system, similar to the one described in [159], has been implemented (Figure 7.4 *red dashed* box). The motivation is to encourage the readers to scrutinize the images carefully: start from a rough decision and then refine their decision for a final score. This would help improve the consistency of their scores. The reported scores at each local VOI for each reader are automatically recorded and saved for analysis. There is no particular order for reporting the four local VOIs, but the readers cannot proceed to the next image until the scores for all local VOIs of the current image are reported. There is a 'VOI response' window next to the image window, showing the reader's progress on the current image, where *light gray* indicates that the score for the corresponding local VOI has been reported, *dark gray* indicates the local VOI that the reader is currently scoring for, and *white* indicates that the score for the local VOI has not been reported yet. A lesion-only VOI, without structured background, is always displayed on the left of the image to remind the readers of the typical appearance of lesions.



Input Images
MF or MC: MF
DBT Geometry AS (24, 60): 24
Np (9, 13, or 21, 31): 13
Image folder: I:\DBT_MMBC_rad\set2
○ Training?

Only after you are done with the 50
the SCORE button will be enabled.

Progress: 2/50

Figure 7.4: Schematic of experimental interface with the three-stage score-reporting system (*red dashed* box)

7.2.2.3 Data collection

All studies are being conducted with a mobile workstation (Dell Inc, Round Rock, TX), and images are displayed on the built-in monitor of resolution 1440 × 900. The window level is chosen to display the variations of background structures [57], and the window width is chosen to be narrow enough to enhance the contrast of the lesions, but not so narrow as to make the image noise distracting. There is no time limit to review each image. Each radiologist needs to read a total of 400 DBT sets in four experimental sessions, one session for each of the four DBT geometries. The images are randomized so that each reader views the images in a different order. In each session, the reader reads 50 MF DBT sets and 50 MC DBT sets. Each DBT set takes on average 15 seconds to read, and the whole experiment takes up to 2 hours for each reader to complete.

7.2.3 Data Analysis

Data analysis of this study focuses on the rank orders of the system configurations for MF versus MC. The observer performance is being evaluated as follows: 1) image-level decisions are measured using the area under the ROC curve ($\mathrm{AUC}_{\mathrm{ROC}}$). The score of the image is defined as the maximum score of the four scores across the four signal locations. 2) location-specific decisions are measured using $\mathrm{AUC}_{\mathrm{AFROC}}$ by plotting the empirical AFROC curve. For each of the chosen decision thresholds, a lesion is counted as correctly-localized if the score at the lesion location is above the threshold, and a lesion is actually present at that location. One false positive is counted if any of the four locations in either a lesion-present or lesion-absent case is a false positive.

As a result, there are a total of four rank orders of the four DBT geometries: (1) image-level rank order for MF cases; (2) image-level rank order for MC cases; (3) location-level rank order for MF cases; and (4) location-level rank order for MC cases. The optimal designs for MF cases and MC cases are assessed by comparing the image-level rank order for MF cases (1) with the image-level rank order for MC cases (2), and comparing location-level rank order for MF cases (3) with the location-level rank order for MC cases (4). To understand if the DBT system design that is optimal for image-level decisions is also optimal for location-level decisions, we compare the image-level rank order for MF cases (1) with the location-level rank order for MF cases (3), and compare image-level rank order for MC cases (2) with the location-level rank order for MC cases (4).

7.3 Preliminary Result

Experimental data utilized in this preliminary analysis were acquired with one board-certified radiologist (radiology faculty, male, 7 years of experience). Figure 7.5 shows the radiologist's detection performance with the four DBT geometries described above. The geometries are shown in the descending order of the FOM (i.e., AUC_{ROC} in Figure 7.5(a) and (b), and $\text{AUC}_{\text{AFROC}}$ in Figure 7.5(c) and (d)), and each geometry is indicated by a unique color. The color patterns in the four subplots of Figure 7.5 are visually different, indicating that the four rank orders of these geometries are not consistent. First, when comparing Figure 7.5(c) with (d), we observe that the geometries $(AS24, Np9)$ (*blue*) and $(AS24, Np13)$ (*cyan*) have the lowest and highest $\text{AUC}_{\text{AFROC}}$, respectively, for detecting MF cases, whereas they have the highest and lowest $\text{AUC}_{\text{AFROC}}$, respectively, for detecting MC cases. These contradicting trends suggest that the rank orders of DBT geometries may be different for detecting MF and for detecting MC cases. Second, when comparing Figure 7.5(a) with (c), we observe that the geometry $(AS24, Np9)$ (*blue*) has the 2^{nd} highest AUC_{ROC} for image-level decisions, but the lowest $\text{AUC}_{\text{AFROC}}$ for location-

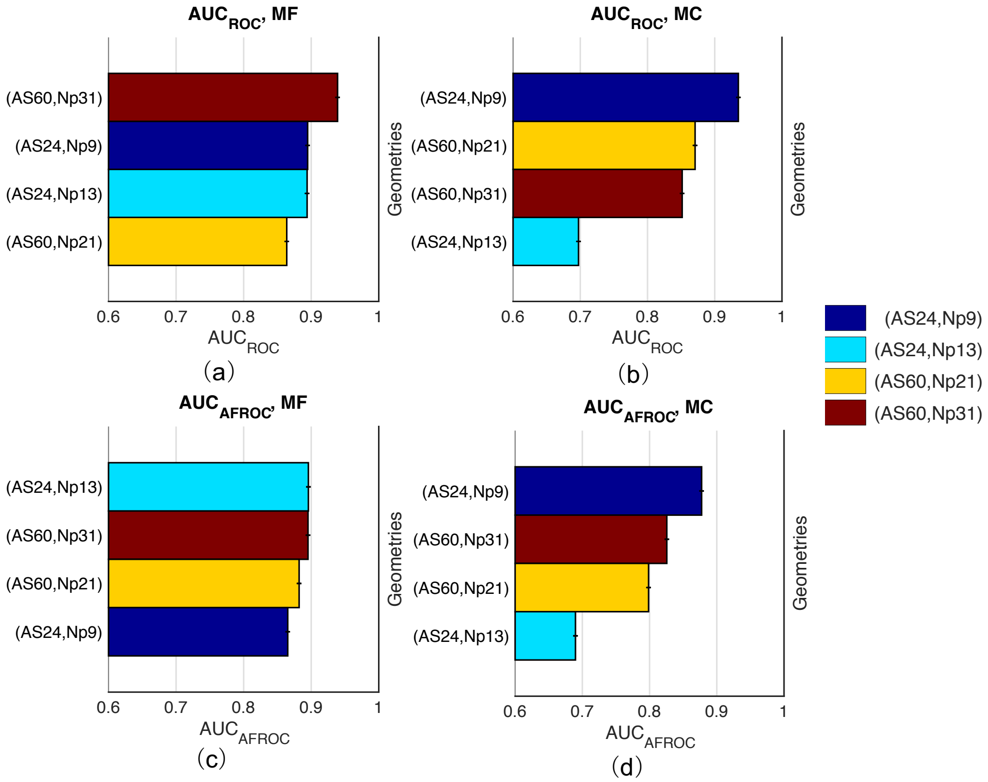


Figure 7.5: Rank orders of the four DBT geometries by: (a) $\mathrm{AUC}_{\mathrm{ROC}}$ in MF cases; (b) $\mathrm{AUC}_{\mathrm{ROC}}$ in MC cases; (c) $\mathrm{AUC}_{\mathrm{AFROC}}$ in MF cases; (d) $\mathrm{AUC}_{\mathrm{AFROC}}$ in MC cases. The performance trends are not consistent when comparing the rank orders: 1) for MF with for MC, and 2) by $\mathrm{AUC}_{\mathrm{ROC}}$ with by $\mathrm{AUC}_{\mathrm{AFROC}}$.

level decisions. In contrast, the geometry $(AS24, Np13)$ $(cyan)$ only has the 3^{rd} highest $\mathrm{AUC}_{\mathrm{ROC}}$ for image-level decisions, but the highest $\mathrm{AUC}_{\mathrm{AFROC}}$ for location-level decisions. These conflicts suggest that the rank orders of DBT geometries may vary as the task of clinical interest changes. Therefore, both qualitative conclusions generally agree with those drawn from our model observer studies (Chapter 6.5).

7.4 Discussion

In this study, we simplify the task of detecting MFMC lesions from DBT reconstructed slices into the task of finding lesions in the extracted 2D VOIs. Such simplification may not reflect how DBT images are interpreted by radiologists in their clinical practice. As the current stack-view presentation approach for volumetric DBT imaging data requires radiologists to scroll back and forth over multiple reconstructed slices in order to find a lesion if any, the real clinical task would be more complicated. Moreover, reading DBT images may be combined with other relevant tasks, such as reading a conventional 2D mammography image or a synthetic 2D mammographic view (e.g., C-view image computed from DBT projections [136, 168]). However, one of the main interests of this study is to examine the potential influence of multiple lesions on DBT system optimization. We are confident that such a simplified detection task, as commonly used in the similar previous studies reported in the literature (e.g., [159, 57, 4], is suitable for our study purpose.

It remains to be seen whether the overall performance trend of human observers would agree with what we observed for our model observer. Here we propose two approaches that may be helpful in future work should there be discrepancies between the human observers and the model observer. 1) The model observer could be modified to include an anthropomorphic module to reflect the same internal noise as the human (e.g., [163, 23, 4])). For example, user scores assigned by humans, instead of the N_l-dimensional binary vectors of 0/1 truth, could serve as $\mathbf{Y}^{\mathrm{L}}$ for generating PLS channels. 2). Intra-observer and inter-observer variability could be prospectively considered for interpreting the results of human observer studies. For example, when assessing the differences between two observers' scores, inter-observer bias and limits of agreement could be calculated [94], which could be further transformed into weighting factors when combining all observers' scores.

Chapter 8

Conclusions and Future Work

In this dissertation, we have advanced two active areas of research in medical imaging. The first area is focused on the model observer design for detecting multiple abnormalities in a single set of medical images. This is motivated by the limitations of existing model observers that they are typically designed to detect at most one signal, and could not address variations across background regions due to the presence of anatomical structures. Our novel multi-signal model observer not only incorporates interactions among signals, but also accounts for variations in local anatomical background by adjusting the decision-making rule accordingly. We have developed novel channelization approaches to estimate efficient channels (PLS and LG channels) that capture significant discriminant image information, such as local background statistics, signal characteristics, and signal prevalence across locations. The evaluation studies with synthetic 2D mammograms have demonstrated that the model observer is able to detect multiple breast lesions, making accurate image-level and location-specific detection decisions with a small number of channels. This shows the potential of the model observer as a broadly applicable tool in medical image quality assessments of realistic clinical situations. This would greatly facilitate the task-based optimization of imaging systems.

The second part of this dissertation aims to optimize DBT image quality for the detection of MFMC breast cancer. As previous DBT system optimization studies are based on unifocal breast cancer scenarios, this is the first series of studies that seek to investigate how the presence of more than one tumor would influence DBT system design. First, we have extended our 2D multi-signal model observer into a 3D ml-CHO model observer that works effectively with volumetric DBT imaging data. The 3D ml-CHO makes detection decisions regarding the presence of lesions upon the overall information from both individual DBT slices and their correlations. In the evaluation studies with simulated DBT images of MFMC cases (i.e., digital breast phantoms with multiple inserted synthetic breast lesions, scanned by a simulated DBT scanner), we have demonstrated that the 3D ml-CHO could achieve high detection performance with a small number of channels (especially with 3D PLS channels). We have also shown that incorporating locally varying anatomical backgrounds and their correlations as in the design of 3D ml-CHO is beneficial for making multi-lesion detection decisions. Second, given these results showing the potential of the 3D ml-CHO as a useful tool in task-based assessment of DBT image quality, we have employed the 3D ml-CHO to identify DBT system geometries that are most effective for the detection of MFMC. We have shown that the consideration of multiple breast lesions may present distinct challenges to DBT system optimization. We have demonstrated that DBT system geometries, especially wide-arc geometries versus narrow-arc geometries, may not be equally effective for detecting MF or MC lesions. We have also shown that the rank orders of the geometries for MF and MC cases and for unifocal cases may be different. Thus, our model observer evaluation studies highlight that the optimal geometry of DBT may vary when the task of clinical interest changes, and a given DBT system geometry may not yield images that are equally informative for detecting MF, MC, and unifocal breast cancers. Overall, all these conclusions are significant and constructive because they could potentially help enhance the clinical values of DBT, and improve management of MFMC.

As this dissertation is the first to investigate the impact of multiple tumors on DBT system optimization, we have brought forward a number of open research questions that need further investigations in the future. First, it is necessary to validate the clinical relevance and significance of our model observer studies. This is consistent with what our ongoing human observer study with experienced breast imaging radiologists (Chapter 7) aims to accomplish. By measuring radiologists' accuracy and reported confidence in detecting synthetic MFMC lesions from simulated DBT images acquired with different imaging geometries, we could rank the chosen DBT systems by the human observer performance, hypothesizing that it would generally agree with the performance trends shown in our model observer studies. On the other hand, the multi-lesion model observer could be improved with an anthropomorphic module to reflect the same internal noise of the human. For example, user scores assigned by humans, instead of the N_l-dimensional binary vectors of 0/1 truth, could serve as $\mathbf{Y}^{\mathrm{L}}$ for generating PLS channels. Second, it is valuable to further explore the realism of simulated MFMC cases. We have only evaluated DBT geometries on cases in which the possible lesion locations are finite and fixed, and lesions always appear in the center of local VOIs. More investigations with varying lesion locations, types and shapes would be valuable. Instead of only working with extracted local VOIs, the 3D ml-CHO could be also extended to consider information from the whole breast, such as the actual distances between lesions, in the decision making process. Third, it has to be noted we have only investigated two key design factors in the design of DBT acquisition process, and we have not exhaustively evaluated all possible combinations of the two parameters. In order to achieve more definitive conclusion on the optimal DBT design for MFMC detection, it would be valuable to conduct more comprehensive studies that evaluate the impacts of other DBT system parameters. For example, variable angular increment and variable dose distribution in the acquisition process are two example factors of interest. Fourth, as our vision in this dissertation is that DBT may offer advantages over MRI in terms of fewer false-positive findings, lower cost, and better accessibility. It would be interesting to directly compare the detectability of MFMC lesions in DBT and MRI images. For instance, multi-modality computational breast phantoms with synthetic breast lesions could be imaged by a simulated DBT scanner, and a simulated MRI scanner. As a result, two sets of images of simulated MFMC cases are generated, and similar model observe studies could be conducted on these images. Detection performance of the multi-lesion model observers could be evaluated for comparing the effectiveness of DBT and MRI for MFMC detection.

Appendices

Appendix A

Chapter 3: Appendix

A.1 Results with realistic signals in multicentric cases

Figure A.1 shows the complete FOM results with realistic signals in multicentric cases. Detailed discussion of the results are in Chapter 3.3.1.

A.2 Results with circular Gaussian signals in multicentric cases

Figure A.2 shows the FOMs with circular Gaussian signals. The general trends of FOMs with multicentric cases as a function of N_c were similar to those reported with multifocal cases in [151].

(i). For image-level SNR, when N_c is small, $\mathbf{T}_{\text{img}}$ PLS channels was superior to $\mathbf{T}_{\text{loc}}$ PLS channels, but $\mathbf{T}_{\text{loc}}$ starts to outperform $\mathbf{T}_{\text{img}}$ when N_c is around 25 (Figure A.2(a)). $\mathbf{T}_{\text{loc}}$ is better than $\mathbf{T}_{\text{img}}$ as $\mathbf{T}_{\text{loc}}$ could be used to generate both image-level and location-specific decisions.

(ii). The performance, measured using $\text{AUC}_{\text{AFROC}}$, of $\mathbf{T}_{\text{loc}}$ PLS channels matches that of $\mathbf{T}_{\text{one\_loc}}$ when N_c per location ≥ 25 (Figure A.2(b)). As $\mathbf{T}_{\text{loc}}$ is universal to all signal locations while $\mathbf{T}_{\text{one\_loc}}$ is specific to one location, there is a huge reduction in total number of channels used (i.e.,

N_c for $\mathbf{T}_{\mathrm{loc}}$ versus $[N_c \times N_l]$ for $\mathbf{T}_{\mathrm{one\_loc}}$). This is especially valuable when N_l is relatively large (e.g., 20).

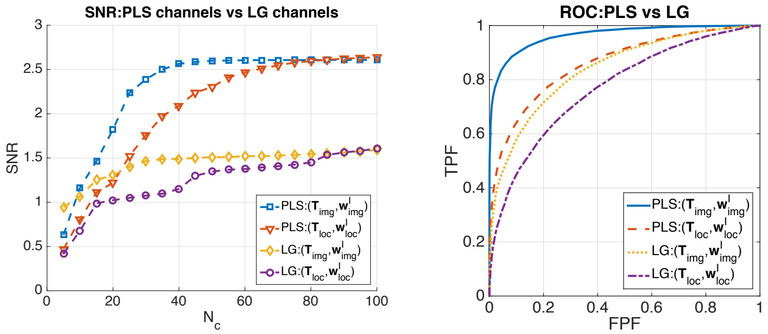
(iii). For both PLS and LG channels, location-specific SNRs were different across the four signal locations, when N_c is sufficiently large (e.g., 50 $\mathbf{T}_{\mathrm{one\_loc}}$ PLS channels per location). This agrees with our assumption that different local background statistics in multicentric cases would lead to different levels of detectability.

(iv). For all three FOMs, the performance trends for PLS channels and modified LG channels were approximately same. The maximum performance with the PLS channels was as good as that with the counterpart LG channels. As LG channels are derived analytically, they are better when the number of training images is limited (i.e., estimated PLS channels are relatively noisy). However, PLS channels are better when realistic signal shapes and complex background statistics are involved in the detection tasks. More results are reported in Chapter 3.3.1.

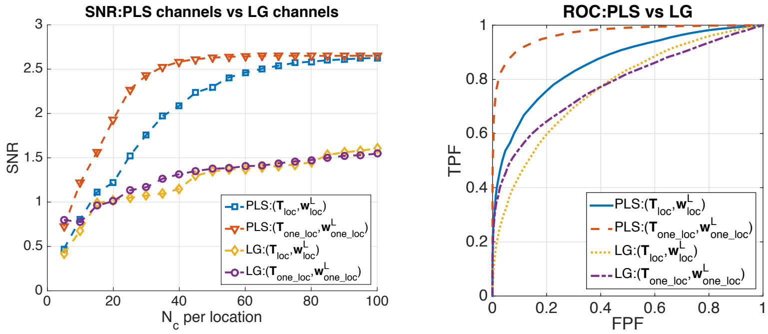
A.3 Interactions among signal locations (LG channels)

Figure A.3 shows the comparisons among the four model observers as discussed in Section 3.3.2. The benchmark LG channels were used.

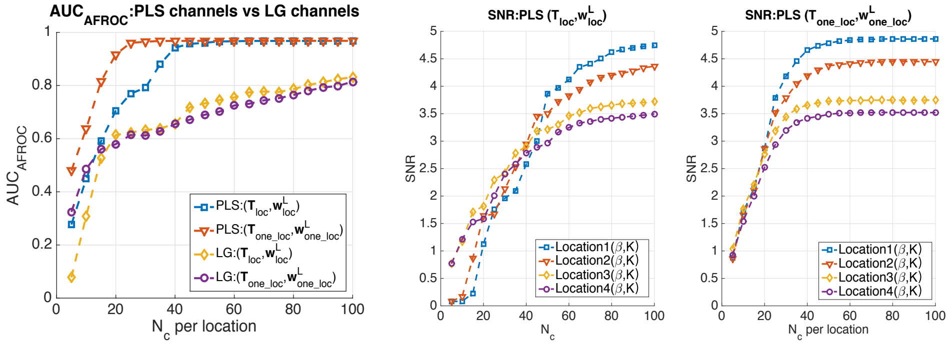
(i). Model 3 and 4 were consistently good when compared to Model 1 and 2 (Figure A.3(a)-(c)).



(a) Image-level SNR ($\mathbf{Y}^{\mathrm{I}}$-trained) (b) ROC curve, $N_c = 25$ ($\mathbf{Y}^{\mathrm{I}}$-trained)

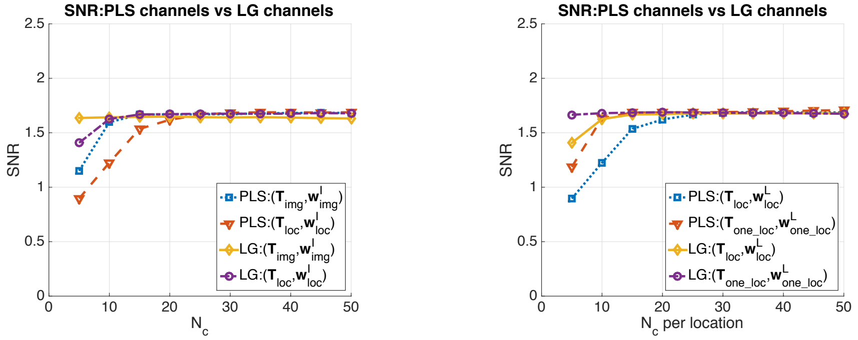


(c) Image-level SNR ($\mathbf{Y}^{\mathrm{L}}$-trained) (d) ROC curve, $N_c = 25$ ($\mathbf{Y}^{\mathrm{L}}$-trained)

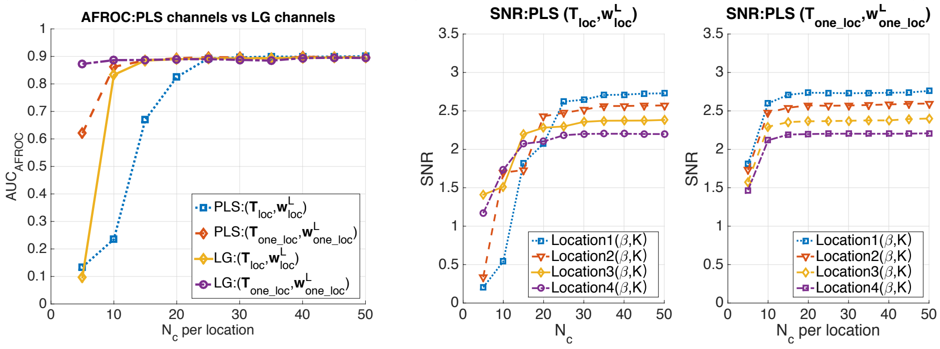


(e) $\mathrm{AUC}_{\mathrm{AFROC}}$ (f) Location-specific SNRs (PLS channels)

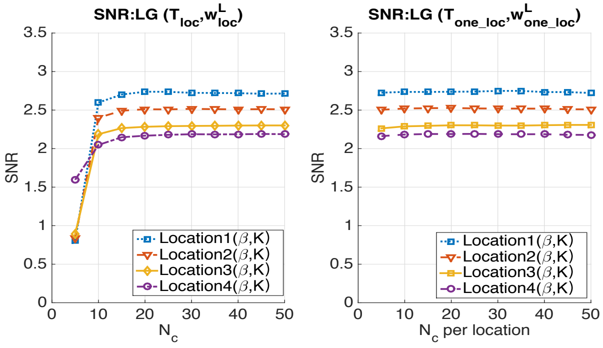
Figure A.1: FOM results with realistic signals in multicentric cases. (a) Image-level decision: SNR from $(\mathbf{T}_{\mathrm{img}}, \mathbf{w}^{\mathrm{I}}_{\mathrm{img}})$ and $(\mathbf{T}_{\mathrm{loc}}, \mathbf{w}^{\mathrm{I}}_{\mathrm{loc}})$, PLS channels versus LG channels; (b) Example ROC curves achieved with $N_c = 25$ from $(\mathbf{T}_{\mathrm{img}}, \mathbf{w}^{\mathrm{I}}_{\mathrm{img}})$ and $(\mathbf{T}_{\mathrm{loc}}, \mathbf{w}^{\mathrm{I}}_{\mathrm{loc}})$: PLS channels versus LG channels; (c) Image-level decision: SNR from $(\mathbf{T}_{\mathrm{loc}}, \mathbf{w}^{\mathrm{L}}_{\mathrm{loc}})$ and $(\mathbf{T}_{\mathrm{one\_loc}}, \mathbf{w}^{\mathrm{L}}_{\mathrm{one\_loc}})$: PLS channels versus LG channels; (d) Example ROC curves achieved with $N_c = 25$ from $(\mathbf{T}_{\mathrm{loc}}, \mathbf{w}^{\mathrm{L}}_{\mathrm{loc}})$ and $(\mathbf{T}_{\mathrm{one\_loc}}, \mathbf{w}^{\mathrm{L}}_{\mathrm{one\_loc}})$: PLS channels versus LG channels; (e) Location-level decision: $\mathrm{AUC}_{\mathrm{AFROC}}$ from $(\mathbf{T}_{\mathrm{loc}}, \mathbf{w}^{\mathrm{L}}_{\mathrm{loc}})$ and $(\mathbf{T}_{\mathrm{one\_loc}}, \mathbf{w}^{\mathrm{L}}_{\mathrm{one\_loc}})$, PLS channels versus LG channels. (f): Location-specific SNRs from PLS $(\mathbf{T}_{\mathrm{loc}}, \mathbf{w}^{\mathrm{L}}_{\mathrm{loc}})$ and PLS $(\mathbf{T}_{\mathrm{one\_loc}}, \mathbf{w}^{\mathrm{L}}_{\mathrm{one\_loc}})$.



(a) Image-level SNR ($\mathbf{Y}^{\mathrm{I}}$-trained) (b) Image-level SNR ($\mathbf{Y}^{\mathrm{L}}$-trained)

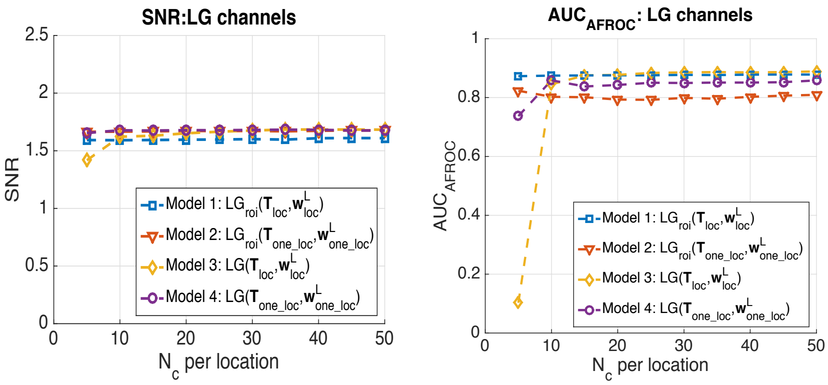


(c) $\mathrm{AUC}_{\mathrm{AFROC}}$ (d) Location-specific SNRs (PLS channels)

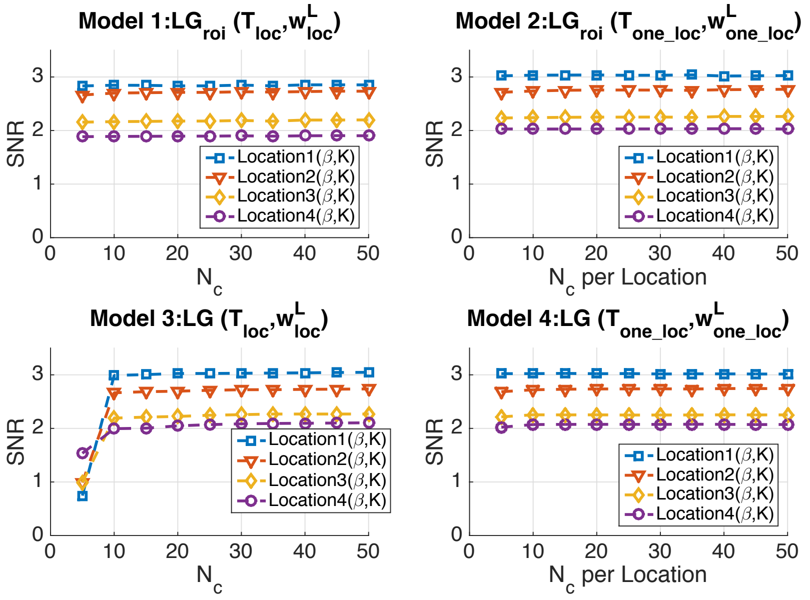


(e) Location-specific SNRs (LG channels)

Figure A.2: FOM results with circular Gaussian signals in multicentric cases.(a) Image-level decision: SNR from $(\mathbf{T}_{\mathrm{img}}, \mathbf{w}^{\mathrm{I}}_{\mathrm{img}})$ and $(\mathbf{T}_{\mathrm{loc}}, \mathbf{w}^{\mathrm{I}}_{\mathrm{loc}})$: PLS channels versus LG channels; (b) Image-level decision: SNR from $(\mathbf{T}_{\mathrm{loc}}, \mathbf{w}^{\mathrm{L}}_{\mathrm{loc}})$ and $(\mathbf{T}_{\mathrm{one\_loc}}, \mathbf{w}^{\mathrm{L}}_{\mathrm{one\_loc}})$: PLS channels versus LG channels; (c) Location-level decision: $\mathrm{AUC}_{\mathrm{AFROC}}$ from $(\mathbf{T}_{\mathrm{loc}}, \mathbf{w}^{\mathrm{L}}_{\mathrm{loc}})$ and $(\mathbf{T}_{\mathrm{one\_loc}}, \mathbf{w}^{\mathrm{L}}_{\mathrm{one\_loc}})$: PLS channels versus LG channels. (d) Location-specific SNRs from $(\mathbf{T}_{\mathrm{loc}}, \mathbf{w}^{\mathrm{L}}_{\mathrm{loc}})$ and $(\mathbf{T}_{\mathrm{one\_loc}}, \mathbf{w}^{\mathrm{L}}_{\mathrm{one\_loc}})$: PLS channels. (e) Location-specific SNRs from $(\mathbf{T}_{\mathrm{loc}}, \mathbf{w}^{\mathrm{L}}_{\mathrm{loc}})$ and $(\mathbf{T}_{\mathrm{one\_loc}}, \mathbf{w}^{\mathrm{L}}_{\mathrm{one\_loc}})$: LG channels.

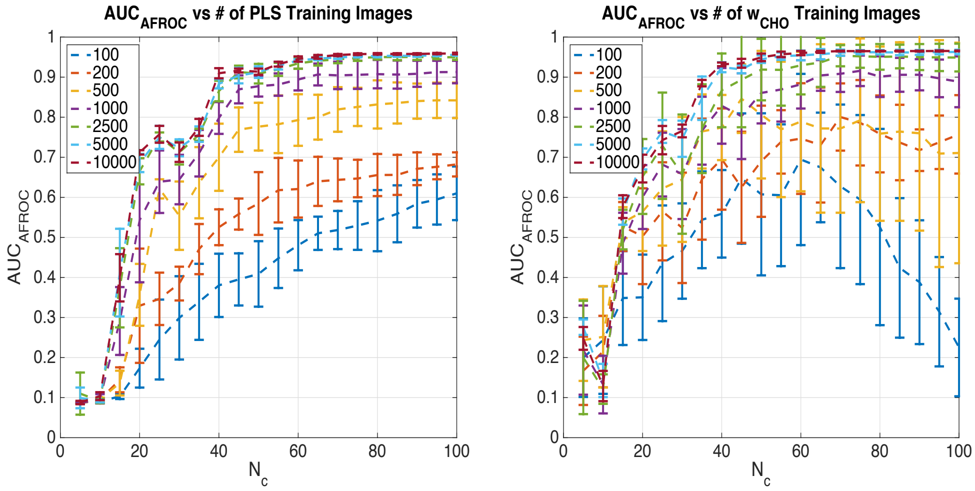


(a) Image-level SNR (b) $\mathrm{AUC}_{\mathrm{AFROC}}$



(c) Location-specific SNRs

Figure A.3: Subfigures (a)-(c) are shown to demonstrate the benefits of incorporating interactions among signal locations in multi-signal detection tasks. The model observers used the modified LG channels. (a) image-level SNR curves show that Model 1 (*blue*) significantly underperforms the other three. (b) $\mathrm{AUC}_{\mathrm{AFROC}}$ curves show that Model 2 (*red*) significantly underperforms the other three. (c) Location-specific SNRs from Model 2 (*top right*) are higher than SNRs from Model 1 (*top left*), but are approximately equal to those from Model 3 (*bottom left*) and 4 (*bottom right*).



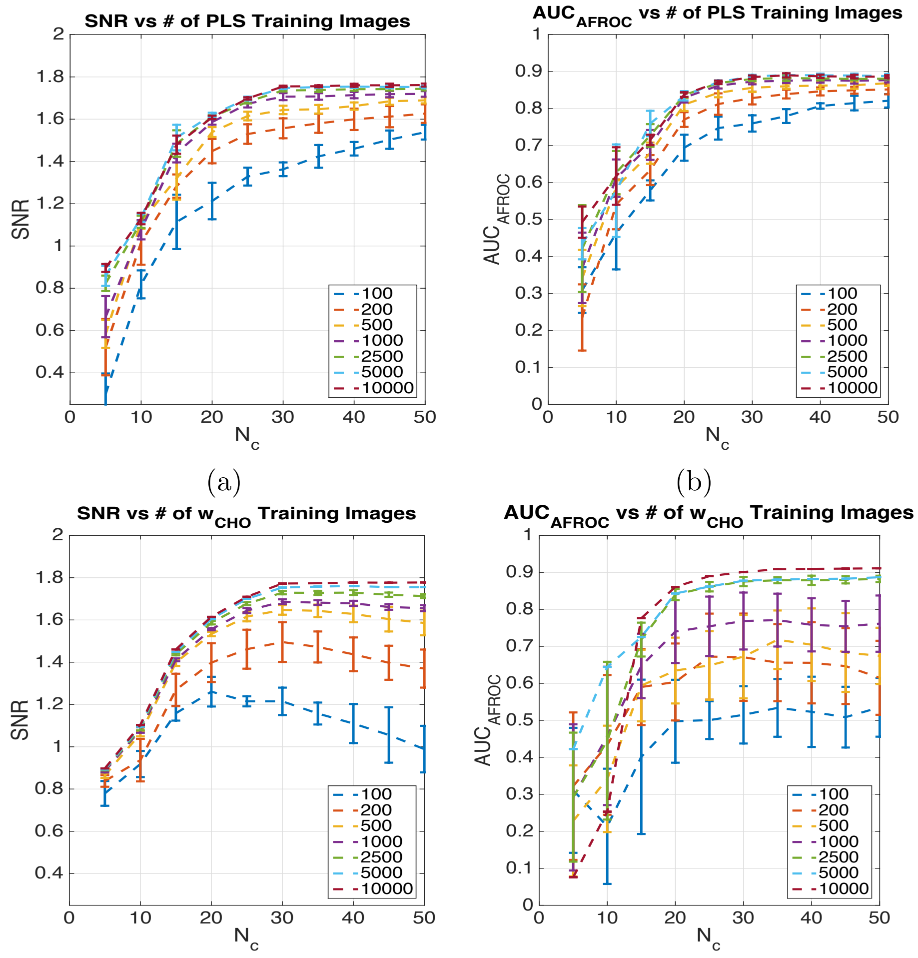
(a) varying number of PLS training (b) varying number of w_{CHO} training

Figure A.4: Impacts of the number of training samples on the performance of the model observers with $\mathbf{T}_{\text{loc}}$ PLS channels. (a)-(b) Plots of $\text{AUC}_{\text{AFROC}}$ as the number of PLS training images, and the number of w_{CHO} training images changes. The performance trend as a function of N_c was similar across different numbers of PLS training images. With smaller number of w_{CHO} training images, $\text{AUC}_{\text{AFROC}}$ was considerably lower with higher variations.

(ii). Model 1 (*blue*), measured by image-level SNR, significantly underperformed the other three models (Figure A.3(a)).

(iii). Model 2 (*red*), measured by $\text{AUC}_{\text{AFROC}}$, significantly underperformed the other three models (Figure A.3(b))

(iv). Model 2 (*top right*) achieved higher location-specific SNRs than Model 1 (*top left*), but approximately equal SNRs as Model 3 and 4 (*bottom*) (Figure A.3(c)).



(a) (b)

(c) (d)

Figure A.5: Impacts of the number of training samples on the performance of the model observers with $\mathbf{T}_{\mathrm{loc}}$ PLS channels. (a)-(b) Plots of image-level SNR and $\mathrm{AUC}_{\mathrm{AFROC}}$ as the number of PLS training images changes. The performance trend as a function of N_c was similar across different numbers of PLS training images. (c)-(d) Plots of image-level SNR and $\mathrm{AUC}_{\mathrm{AFROC}}$ as the number of $\mathrm{w}_{\mathrm{CHO}}$ training images changes. With smaller number of $\mathrm{w}_{\mathrm{CHO}}$ training images, image-level SNR degraded significantly at large N_c, and $\mathrm{AUC}_{\mathrm{AFROC}}$ was considerably lower with higher variations.

A.4 Sample size considerations (realistic signals)

Figure A.4 shows the impacts of sample sizes on $\mathrm{AUC}_{\mathrm{AFROC}}$. Irregularly shaped signals were used. Despite the notable differences in the absolute performance, the trend of $\mathrm{AUC}_{\mathrm{AFROC}}$ (Figure A.4(a)) as a function of N_c was similar across different numbers of PLS training images. However, for small numbers of $\mathrm{w}_{\mathrm{CHO}}$ training images, $\mathrm{AUC}_{\mathrm{AFROC}}$ at a large N_c was substantially lower with a larger variation (Figure A.4(b)).

A.5 Sample size considerations (circular Gaussian signals)

Figure A.5 shows the impacts of sample sizes on the detection performance. Circular Gaussian signals were used.

A.5.1 The number of PLS training images

We fixed the number of $\mathrm{w}_{\mathrm{CHO}}$ training images and testing images at 5000, and varied the number of PLS training images to one of $\{100, 200, 500, 1000, 2500, 5000, 10000\}$. Each setting was repeated for five times, and the templates $\mathbf{w}$ were sufficiently trained for each setting. The trends of FOMs (Figure A.5(a)-(b)) as a function of N_c were similar across different numbers of PLS training images. Hence, PLS channels, estimated from a limited number of training images, could still capture useful information for predicting performance trend in multi-signal detection tasks, even if the quality of the channels may not as good as they could possibly be.

A.5.2 The number of $\mathbf{w}_{\mathbf{CHO}}$ training images

Similarly, we fixed the number of PLS training images and testing images at 5000, and varied the number of $\mathrm{w}_{\mathrm{CHO}}$ training images to one of $\{100, 200, 500, 1000, 2500, 5000, 10000\}$. For small numbers of $\mathrm{w}_{\mathrm{CHO}}$ training numbers (e.g., 100, 200), it can be seen that the image-level SNR (Figure A.5(c)) significantly degraded with a large N_c, and $\mathrm{AUC}_{\mathrm{AFROC}}$ (Figure A.5(d)) was substantially lower with a larger variation. This may be caused by inaccurate, and/or instable estimation of the channelized covariance, when such a small number of training images were used. Given the quality of PLS channels was consistently high, this suggests that the sample size of $\mathrm{w}_{\mathrm{CHO}}$ training may have a larger impact on the performance. Thus, if the total sample size is limited, a larger fraction of the images need be used to train the template $\mathbf{w}$.

Appendix B

Chapter 6: Appendix

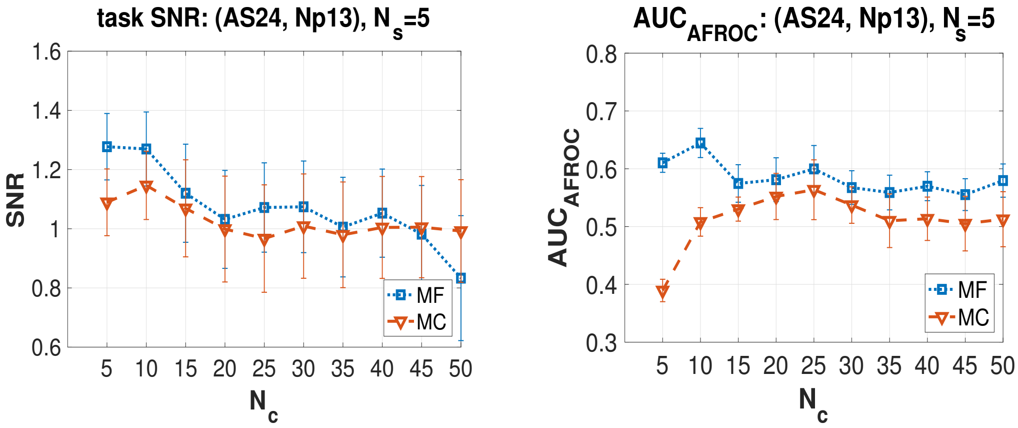
B.1 Reconstruction Algorithms

Though image acquisition decides what and how much information is captured, there are different types of DBT reconstruction algorithms [128] (e.g., linear and non-linear methods, iterative reconstruction) that impact resulting DBT image properties (e.g., spatial resolution, noise level, contrast, and artifacts). In Chapter 6, all the images were reconstructed by the FBP method with a Hann filter. To verify if the choice of reconstruction algorithm would affect the DBT system optimization for MFMC detection, we briefly explored other alternative algorithms, including the simultaneous algebraic reconstruction technique (SART) (e.g., [84, 164]), the maximum-likelihood method (e.g., [119]), and the total variation minimization method (e.g., [134, 48]). Figure B.1 and Figure B.2, as the counterparts of Figure 6.4 and Figure 6.6 respectively, show the preliminary results with SART. The step size of the iterative reconstruction was manually tuned to achieve a fast convergence of the cost function while still being able to reach a stable solution. The iteration started with an initial guess obtained with Back Projection, and the maximum number of iterations was set to 20. The same procedures of 3D ml-CHO training and testing as described in Chapter 6.2.3 were used.

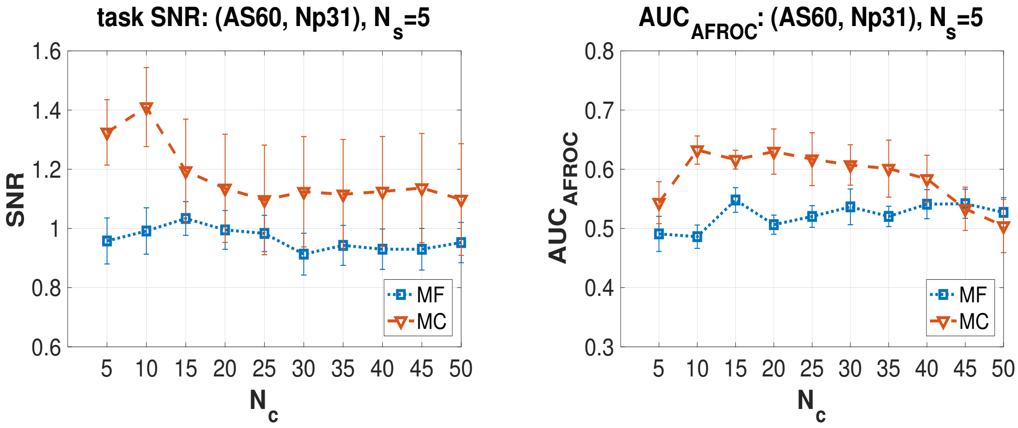
(i). Similar to the results in Chapter 6.3.1, for the geometry $(AS24, Np13)$, both the task SNR (Figure B.1(a)) and $\text{AUC}_{\text{AFROC}}$ (Figure B.1(b)) for MF cases (*blue*) was higher than those for MC cases (*red*). This suggests that for the narrow-arc geometry, MC cases may be more difficult than MF cases to detect accurately. In contrast, for the geometry $(AS60, Np31)$, both the task SNR (Figure B.1(c)) and $\text{AUC}_{\text{AFROC}}$ (Figure B.1(d)) for MF cases (*blue*) was lower than those for MC cases (*red*). This suggests that for wide-arc geometry, MF cases may be more difficult than MC cases.

(ii). Similar to the results in Chapter 6.3.2, for both wide-arc geometry $(AS60)$ and narrow arc-geometry $(AS24)$, changing the number of projections Np changed the trend of detection performance. For example, in Figure B.2(a) of task SNR, the geometry $(AS24, Np13)$ (*red*) has a higher SNR than the geometry $(AS24, Np9)$ (*blue*) when $N_s = 5$, but a lower SNR when $N_s = 11$. Moreover, the rank orders of the four geometries may be not the same by task SNR or $\text{AUC}_{\text{AFROC}}$. For example, for $N_s \geq 7$, the geometry $(AS60, Np21)$ (*yellow*) was significantly better than the geometry $(AS24, Np13)$ (*blue*) by task SNR (Figure B.2(a)), but worse by $\text{AUC}_{\text{AFROC}}$ (Figure B.2(b)).

In conclusion, these results demonstrate that the FBP-based comparisons of the geometries could be generalized to DBT systems with a different reconstruction algorithm.

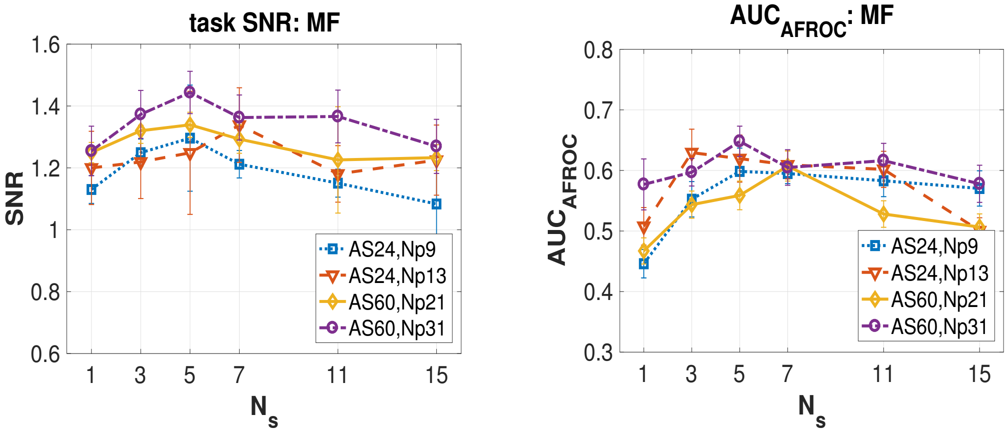


(a) task SNR $(AS24, Np13)$ (b) AUC$_{\text{AFROC}}$ $(AS24, Np13)$



(c) task SNR $(AS60, Np31)$ (d) AUC$_{\text{AFROC}}$ $(AS60, Np31)$

Figure B.1: FOM comparisons between MF cases and MC cases when SART was used for reconstruction. 5 reconstructed slices were used by the 3D ml-CHO. (a)-(b): The DBT geometry $(AS24, Np13)$ may be more effective for detecting lesions in MF cases (blue) than MC cases (red). (c)-(d): The DBT geometry $(AS60, Np31)$ may be more effective for MC cases (red) than MF cases $(blue)$.



(a) task SNR for MF cases (b) AUC$_{\mathrm{AFROC}}$ for MF cases

Figure B.2: (a)-(b): FOM comparisons in MF cases across the four DBT geometries when SART was used for reconstruction. 15 PLS channels (i.e., $N_c = 15$) were used by each of the differently trained four model observers. (c)-(d): the counterpart FOM comparisons in MC cases. For both wide- and narrow arc-geometries ($AS60$ and $AS24$), changing Np changed the trends of detection performance in MF cases, but not in MC cases. The rank orders of the four geometries may not be the same by the task SNR or AUC$_{\mathrm{AFROC}}$ in MF cases or MC cases.

Bibliography

[1] Craig K Abbey and Harrison H Barrett. Human-and model-observer performance in ramp-spectrum noise: effects of regularization and object variability. *JOSA A*, 18(3):473–488, 2001.

[2] Craig K Abbey and John M Boone. An ideal observer for a model of x-ray imaging in breast parenchymal tissue. In *International Workshop on Digital Mammography*, pages 393–400. Springer, 2008.

[3] Aleodor A Andea, Tracie Wallis, Lisa A Newman, David Bouwman, Jyotirmoy Dey, and Daniel W Visscher. Pathologic analysis of tumor size and lymph node status in multifocal/multicentric breast carcinoma. *Cancer*, 94(5):1383–1390, 2002.

[4] Ali RN Avanaki, Kathryn S Espig, Tom RL Kimpe, and Andrew DA Maidment. On anthropomorphic decision making in a model observer. In *SPIE Medical Imaging*, pages 941610–941610. International Society for Optics and Photonics, 2015.

[5] Predrag R Bakic, Michael Albert, Dragana Brzakovic, and Andrew DA Maidment. Mammogram synthesis using a 3d simulation. ii. evaluation of synthetic mammogram texture. *Medical Physics*, 29(9):2140–2151, 2002.

[6] Predrag R Bakic, Cuiping Zhang, and Andrew DA Maidment. Development and characterization of an anthropomorphic breast software phantom based upon region-growing algorithm. *Medical Physics*, 38(6):3165–3176, 2011.

[7] Kheya Banerjee, Anando Sen, Dev P. Chakraborty, and Howard C. Gifford. A visual-search model observer for multitarget volumetric spect images. In *The 13th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, pages 114–121. Fully 3D 2015, 2015.

[8] Jing Bao, Ke-Da Yu, Yi-Zhou Jiang, Zhi-Ming Shao, and Gen-Hong Di. The effect of laterality and primary tumor site on cancer-specific mortality in breast cancer: A seer population-based study. *PloS one*, 9(4):e94815, 2014.

[9] Harrison H Barrett, Craig K Abbey, and Eric Clarkson. Objective assessment of image quality. iii. roc metrics, ideal observers, and likelihood-generating functions. *JOSA A*, 15(6):1520–1535, 1998.

[10] Harrison H Barrett and Kyle J Myers. Foundations of image science. wiler-interscience. *John Wiley & Sons, Inc., New York, New York, USA*, 1(4):17, 2004.

[11] Harrison H Barrett, Jie Yao, Jannick P Rolland, and Kyle J Myers. Model observers for assessment of image quality. *Proceedings of the National Academy of Sciences*, 90(21):9758–9765, 1993.

[12] Kevin S Berbaum, George Y El-Khoury, Kenjirou Ohashi, Kevin M Schartz, Robert T Caldwell, Mark Madsen, and Edmund A Franken. Satisfaction of search in multitrauma patients: severity of detected fractures. *Academic radiology*, 14(6):711–722, 2007.

[13] Kevin S Berbaum, Edmund A Franken Jr, Donald D Dorfman, Seyed A Rooholamini, Mary H Kathol, Thomas J Barloon, Frank M Behlke, Yutaka Sato, Charles H Lu, George Y El-khoury, et al. Satisfaction of search in diagnostic radiology. *Investigative radiology*, 25(2):133–140, 1990.

[14] Wendie A Berg and Patricia L Gilbreath. Multicentric and multifocal cancer: Whole-breast us in preoperative evaluation 1. *Radiology*, 214(1):59–66, 2000.

[15] Kirby I Bland, Herman R Menck, Carol EH Scott-Conner, Monica Morrow, David J Winchester, and David P Winchester. The national cancer data base 10-year survey of breast carcinoma treatment at hospitals in the united states. *Cancer*, 83(6):1262–1273, 1998.

[16] Francois Bochud, Craig Abbey, and Miguel Eckstein. Statistical texture synthesis of mammographic images with super-blob lumpy backgrounds. *Optics Express*, 4(1):33–42, 1999.

[17] François O Bochud, Jean-François Valley, Francis R Verdun, Christian Hessler, and Pierre Schnyder. Estimation of the noisy component of anatomical backgrounds. *Medical Physics*, 26(7):1365–1370, 1999.

[18] John M Boone, Karen K Lindfors, Virgil N Cooper, and J Anthony Seibert. Scatter/primary in mammography: Comprehensive results. *Medical Physics*, 27(10):2408–2416, 2000.

[19] Lucas R Borges, Helder CR de Oliveira, Polyana F Nunes, Predrag R Bakic, Andrew DA Maidment, and Marcelo AC Vieira. Method for simulating dose reduction in digital mammography using the anscombe transformation. *Medical Physics*, 43(6):2704–2714, 2016.

[20] R Bowditch. Patterns found in false negative cervical cytology. *Cytoletter*, 3:22–25, 1996.

[21] Anna Bozzini, Giuseppe Renne, Lorenza Meneghetti, Giuseppe Bandi, Gabriela Santos, Anna Rita Vento, Simona Menna, Stefania Andrighetto, Giuseppe Viale, Enrico Cassano, et al. Sensitivity of imaging for multifocal-multicentric breast carcinoma. *BMC cancer*, 8(1):275, 2008.

[22] Kathleen R Brandt, Daniel A Craig, Tanya L Hoskins, Tara L Henrichsen, Emily C Bendel, Stephanie R Brandt, and Jay Mandrekar. Can digital breast tomosynthesis replace conventional diagnostic mammography views for screening recalls without calcifications? a comparison study in a simulated clinical setting. *American Journal of Roentgenology*, 200(2):291–298, 2013.

[23] Jovan G Brankov. Optimization of the internal noise models for channelized hotelling observer. In *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*, pages 1788–1791. IEEE, 2011.

[24] Federico Buggi, Annalisa Curcio, Fabio Falcini, and Secondo Folli. Multicentric/multifocal breast cancer: overview, biology, and therapy. In *Cell and Molecular Biology of Breast Cancer*, pages 29–42. Springer, 2013.

[25] Arthur E Burgess. Visual perception studies and observer models in medical imaging. 41(6):419–436, 2011.

[26] Arthur E Burgess, Francine L Jacobson, and Philip F Judy. Human observer detection experiments with mammograms and power-law noise. *Medical Physics*, 28(4):419–437, 2001.

[27] Arthur E Burgess and Philip F Judy. Signal detection in power-law noise: effect of spectrum exponents. *JOSA A*, 24(12):B52–B60, 2007.

[28] Cyril Castella, Karen Kinkel, François Descombes, Miguel P Eckstein, Pierre-Edouard Sottas, Francis R Verdun, and François O Bochud. Mammographic texture synthesis: second-generation clustered lumpy backgrounds using a genetic algorithm. *Optics Express*, 16(11):7595–7607, 2008.

[29] Björn Cederström and Erik Fredenberg. The influence of anatomical noise on optimal beam quality in mammography. *Medical Physics*, 41(12):121903, 2014.

[30] Dev P Chakraborty. A brief history of free-response receiver operating

characteristic paradigm data analysis. *Academic Radiology*, 20(7):915–919, 2013.

[31] Heang-Ping Chan, Mitchell M Goodsitt, Mark A Helvie, Scott Zelakiewicz, Andrea Schmitz, Mitra Noroozian, Chintana Paramagul, Marilyn A Roubidoux, Alexis V Nees, Colleen H Neal, et al. Digital breast tomosynthesis: Observer performance of clustered microcalcification detection on breast phantom images acquired with an experimental system using variable scan angles, angular increments, and number of projection views. *Radiology*, 273(3):675–685, 2014.

[32] Amarpreet S Chawla, Joseph Y Lo, Jay A Baker, and Ehsan Samei. Optimized image acquisition for breast tomosynthesis in projection and reconstruction space. *Medical Physics*, 36(11):4859–4869, 2009.

[33] Baiyu Chen, Lifeng Yu, Shuai Leng, James Kofler, Christopher Favazza, Thomas Vrieze, and Cynthia McCollough. Predicting detection performance with model observers: Fourier domain or spatial domain? In *SPIE Medical Imaging*, pages 978326–978326. International Society for Optics and Photonics, 2016.

[34] Lin Chen, Craig K Abbey, and John M Boone. Association between power law coefficients of the anatomical noise power spectrum and lesion detectability in breast imaging modalities. *Physics in Medicine and Biology*, 58(6):1663, 2013.

[35] Mu Chen, James E Bowsher, Alan H Baydush, Karen L Gilland, David M DeLong, and Ronald J Jaszczak. Using the hotelling observer on multislice and multiview simulated spect myocardial images. *IEEE Transactions on Nuclear Science*, 49(3):661–667, 2002.

[36] Xinyuan Chen, Xiaolin Gong, Christian G Graff, Maira Santana, Gregory M Sturgeon, Thomas J Sauer, Rongping Zeng, Stephen J Glick, and Joseph Y Lo. High-resolution, anthropomorphic, computational breast phantom: fusion of rule-based structures with patient-based anatomy. In *SPIE Medical Imaging*, pages 101321W–101321W. International Society for Optics and Photonics, 2017.

[37] Paola Clauser, Luca A Carbonaro, Martina Pancot, Rossano Girometti, Massimo Bazzocchi, Chiara Zuiani, and Francesco Sardanelli. Additional findings at preoperative breast mri: the value of second-look digital breast tomosynthesis. *European radiology*, 25(10):2830–2839, 2015.

[38] William G Cochran. *Sampling techniques.* John Wiley & Sons, 2007.

[39] NJ Coombs and J Boyages. Multifocal and multicentric breast cancer: does each focus matter? *Journal of Clinical Oncology*, 23(30):7497–7502, 2005.

[40] Mini Das, Howard C Gifford, Stephen J Glick, et al. Evaluation of a variable dose acquisition technique for microcalcification and mass detection in digital breast tomosynthesis. *Medical Physics*, 36(6):1976–1984, 2009.

[41] Peter J Dawson, Paul A Baekey, and Robert A Clark. Mechanisms of multifocal breast cancer: an immunocytochemical study. *Human pathology*, 26(9):965–969, 1995.

[42] Sijmen De Jong. Simpls: an alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, 18(3):251–263, 1993.

[43] Richard M DeMay. Common problems in papanicolaou smear interpretation. *Archives of pathology & laboratory medicine*, 121(3):229, 1997.

[44] Stephen B Edge and Carolyn C Compton. The american joint committee on cancer: the 7th edition of the ajcc cancer staging manual and the future of tnm. *Annals of Surgical Oncology*, 17(6):1471–1474, 2010.

[45] Heba G El-Sheredy, Sanaa A El-Benhawy, Khaled Matrawy, Rabie Ramadan, and Yasser Hamed. Multifocal/multicentric versus unifocal breast cancer: What is the difference? *Middle East Journal of Cancer*, 7(2):69–78, 2016.

[46] Emma Engstrom, Ingrid Reiser, and Robert Nishikawa. Comparison of power spectra for tomosynthesis projections and reconstructed images. *Medical Physics*, 36(5):1753–1758, 2009.

[47] David W Erickson, Jered R Wells, Gregory M Sturgeon, Ehsan Samei, James T Dobbins, W Paul Segars, and Joseph Y Lo. Population of 224

realistic human subject-based computational breast phantoms. *Medical Physics*, 43(1):23–32, 2016.

[48] Metin Ertas, Isa Yildirim, Mustafa Kamasak, and Aydin Akan. Digital breast tomosynthesis image reconstruction using 2d and 3d total variation minimization. *Biomedical engineering online*, 12(1):112, 2013.

[49] Steve Si Jia Feng, Carl J D'Orsi, Mary S Newell, Rebecca L Seidel, Bhavika Patel, and Ioannis Sechopoulos. X-ray scatter correction in breast tomosynthesis with a precomputed scatter map library. *Medical Physics*, 41(3), 2014.

[50] Steve Si Jia Feng and Ioannis Sechopoulos. Clinical digital breast tomosynthesis system: dosimetric characterization. *Radiology*, 263(1):35–42, 2012.

[51] K Fezzani, J Sage, I Fitton, L Hadid, A Moussier, N Pierrat, A Martineau, and C Etard. Digital breast tomosynthesis: Dose and image quality characterization. *Physica Medica*, 32:276, 2016.

[52] Uwe Fischer, Lars Kopka, and Eckhardt Grabbe. Breast carcinoma: Effect of preoperative contrast-enhanced mr imaging on the therapeutic approach 1. *Radiology*, 213(3):881–888, 1999.

[53] Mathias S Fleck, Ehsan Samei, and Stephen R Mitroff. Generalized satisfaction of search: Adverse influences on dual-target search accuracy. *Journal of Experimental Psychology: Applied*, 16(1):60, 2010.

[54] Brandon D Gallas and Harrison H Barrett. Validating the use of channels to estimate the ideal linear observer. *JOSA A*, 20(9):1725–1738, 2003.

[55] Howard C Gifford, Zhihua Liang, and Mini Das. Visual-search observers for assessing tomographic x-ray image quality. *Medical Physics*, 43(3):1563–1575, 2016.

[56] Xing Gong, Stephen J Glick, Bob Liu, Aruna A Vedula, and Samta Thacker. A computer simulation study comparing lesion detection accuracy with digital mammography, breast tomosynthesis, and cone-beam ct breast imaging. *Medical Physics*, 33(4):1041–1052, 2006.

[57] Mitchell M Goodsitt, Heang-Ping Chan, Andrea Schmitz, Scott Zelakiewicz, Santosh Telang, Lubomir Hadjiiski, Kuanwong Watcharotone, Mark A Helvie, Chintana Paramagul, Colleen Neal, et al. Digital breast tomosynthesis: studies of the effects of acquisition geometry on contrast-to-noise ratio and observer preference of low-contrast objects in breast phantom images. *Physics in Medicine and Biology*, 59(19):5883, 2014.

[58] Christian G Graff and Kyle J Myers. The ideal observer objective assessment metric for magnetic resonance imaging. In *Biennial International Conference on Information Processing in Medical Imaging*, pages 760–771. Springer, 2011.

[59] Christiane M Hakim, Denise M Chough, Marie A Ganott, Jules H Sumkin, Margarita L Zuley, and David Gur. Digital breast tomosynthesis in the

diagnostic environment: a subjective side-by-side review. *American Journal of Roentgenology*, 195(2):W172–W176, 2010.

[60] SE Harms, DP Flamig, KL Hesley, MD Meiches, RA Jensen, WP Evans, DA Savino, and RV Wells. Mr imaging of the breast with rotating delivery of excitation off resonance: clinical experience with pathologic correlation. *Radiology*, 187(2):493–501, 1993.

[61] Xin He and Subok Park. Model observers in medical imaging research. *Theranostics*, 3(10):774, 2013.

[62] Anne P Hillstrom. Repetition effects in visual search. *Perception & Psychophysics*, 62(4):800–817, 2000.

[63] Alexander Hlawatsch, Andrea Teifke, Marcus Schmidt, and Manfred Thelen. Preoperative assessment of breast cancer: sonography versus mr imaging. *American Journal of Roentgenology*, 179(6):1493–1501, 2002.

[64] Nehmat Houssami, Stefano Ciatto, Petra Macaskill, Sarah J Lord, Ruth M Warren, J Michael Dixon, and Les Irwig. Accuracy and surgical impact of magnetic resonance imaging in breast cancer staging: systematic review and meta-analysis in detection of multifocal and multicentric cancer. *Journal of Clinical Oncology*, 26(19):3248–3258, 2008.

[65] Christina ML Hsu, Mark L Palmeri, W Paul Segars, Alexander I Veress, and James T Dobbins. Generation of a suite of 3d computer-generated

breast phantoms from a limited set of human subject data. *Medical Physics*, 40(4), 2013.

[66] Yue-Houng Hu, Bo Zhao, and Wei Zhao. Image artifacts in digital breast tomosynthesis: Investigation of the effects of system geometry and reconstruction parameters using a linear system approach. *Medical Physics*, 35(12):5242–5252, 2008.

[67] Walter Huda, Kent M Ogden, Ernest M Scalzetti, David R Dance, and Elizabeth A Bertrand. How do lesion size and random noise affect detection performance in digital mammography? *Academic radiology*, 13(11):1355–1366, 2006.

[68] Lynda Ikejimba, Joseph Y Lo, Yicheng Chen, Nadia Oberhofer, Nooshin Kiarashi, and Ehsan Samei. A quantitative metrology for performance characterization of five breast tomosynthesis systems based on an anthropomorphic phantom. *Medical Physics*, 43(4):1627–1638, 2016.

[69] Judy R James, William Pavlicek, James A Hanson, Thomas F Boltz, and Bhavika K Patel. Breast radiation dose with cesm compared with 2d ffdm and 3d tomosynthesis mammography. *American Journal of Roentgenology*, 208(2):362–372, 2017.

[70] Nooshin Kiarashi, Yuan Lin, William P Segars, Sujata V Ghate, Lynda Ikejimba, Baiyu Chen, Joseph Y Lo, James T Dobbins, Loren W Nolte, and Ehsan Samei. Development of a dynamic 4d anthropomorphic

breast phantom for contrast-based breast imaging. In *SPIE Medical Imaging*, pages 83130C–83130C. International Society for Optics and Photonics, 2012.

[71] Nooshin Kiarashi, Loren W Nolte, Joseph Y Lo, W Paul Segars, Sujata V Ghate, Justin B Solomon, and Ehsan Samei. Impact of breast structure on lesion detection in breast tomosynthesis, a simulation study. *Journal of Medical Imaging*, 3(3):035504–035504, 2016.

[72] Jae-Seung Kim, Paul E Kinahan, Carole Lartizien, Claude Comtat, and Thomas K Lewellen. A comparison of planar versus volumetric numerical observers for detection task performance in whole-body pet imaging. *IEEE Transactions on Nuclear Science*, 51(1):34–40, 2004.

[73] Elizabeth A Krupinski. Current perspectives in medical image perception. *Attention, Perception, & Psychophysics*, 72(5):1205–1217, 2010.

[74] Elizabeth A Krupinski. The role of perception in imaging: past and future. In *Seminars in Nuclear Medicine*, volume 41, pages 392–400. Elsevier, 2011.

[75] Harold L Kundel, Calvin F Nodine, Emily F Conant, and Susan P Weinstein. Holistic component of image perception in mammogram interpretation: gaze-tracking study 1. *Radiology*, 242(2):396–402, 2007.

[76] Harold L Kundel, Calvin F Nodine, and Elizabeth A Krupinski. Searching for lung nodules: Visual dwell indicates locations of false-positive and

false-negative decisions. *Investigative Radiology*, 24(6):472–478, 1989.

[77] Harold L Kundel, Calvin F Nodine, Elizabeth A Krupinski, and Claudia Mello-Thoms. Using gaze-tracking data and mixture distribution analysis to support a holistic model for the detection of cancers on mammograms. *Academic Radiology*, 15(7):881–886, 2008.

[78] Beverly A Lau, Mini Das, and Howard C Gifford. Towards visual-search model observers for mass detection in breast tomosynthesis. In *SPIE Medical Imaging*, pages 86680X–86680X. International Society for Optics and Photonics, 2013.

[79] Beverly A Lau, Ingrid Reiser, Robert M Nishikawa, and Predrag R Bakic. A statistically defined anthropomorphic software breast phantom. *Medical Physics*, 39(6):3375–3385, 2012.

[80] Andrew HS Lee. Why is carcinoma of the breast more frequent in the upper outer quadrant? a case series based on needle core biopsy diagnoses. *The Breast*, 14(2):151–152, 2005.

[81] Changwoo Lee, Jongduk Baek, and Subok Park. Investigation on location-dependent detectability of a small mass for digital breast tomosynthesis evaluation. In *SPIE Medical Imaging*, pages 97870V–97870V. International Society for Optics and Photonics, 2016.

[82] Sandy C Lee, Payal A Jain, Samir C Jethwa, Debu Tripathy, and Mary W Yamashita. Radiologist's role in breast cancer staging: providing key

information for clinicians. *Radiographics*, 34(2):330–342, 2014.

[83] Bob Liu, Tao Wu, Richard H Moore, and Daniel B Kopans. Monte carlo simulation of x-ray scatter based on patient model from digital breast tomosynthesis. In *Medical Imaging*, pages 61421N–61421N. International Society for Optics and Photonics, 2006.

[84] Yao Lu, Heang-Ping Chan, Jun Wei, Mitch Goodsitt, Paul L Carson, Lubomir Hadjiiski, Andrea Schmitz, Jeffrey W Eberhard, and Bernhard EH Claus. Image quality of microcalcifications in digital breast tomosynthesis: Effects of projection-view distributions. *Medical Physics*, 38(10):5703–5712, 2011.

[85] James G Mainprize, Albert H Tyson, and Martin J Yaffe. The relationship between anatomic noise and volumetric breast density for digital mammography. *Medical Physics*, 39(8):4660–4668, 2012.

[86] James G Mainprize and Martin J Yaffe. A breast density-dependent power-law model for digital mammography. In *International Workshop on Digital Mammography*, pages 761–768. Springer, 2012.

[87] Serghei Malkov, Lin Ma, Karla Kerlikowske, Jeff Wang, Steve Cummings, and John Shepherd. Comparison of subregional breast density with whole breast density. In *Digital Mammography*, pages 402–407. Springer, 2010.

[88] De May. Rm the art & science of cytopathology, aspiration cytology, 1996.

[89] J Mercier, F Kwiatkowski, C Abrial, V Boussion, V Dieu-de Fraissinette, W Marraoui, V Petitcolin-Bidet, and S Lemery. The role of tomosynthesis in breast cancer staging in 75 patients. *Diagnostic and interventional imaging*, 96(1):27–35, 2015.

[90] Kathrine G Metheany, Craig K Abbey, Nathan Packard, and John M Boone. Characterizing anatomical variability in breast ct images. *Medical Physics*, 35(10):4685–4694, 2008.

[91] Lavinia P Middleton, George Vlastos, Nadeem Q Mirza, S Eva Singletary, and Aysegul A Sahin. Multicentric mammary carcinoma. *Cancer*, 94(7):1910–1916, 2002.

[92] I Millet, E Pages, D Hoa, S Merigeaud, F Curros Doyon, X Prat, and P Taourel. Pearls and pitfalls in breast mri. *The British journal of radiology*, 2014.

[93] H Mumtaz, MA Hall-Craggs, T Davidson, K Walmsley, W Thurell, MW Kissin, and I Taylor. Staging of symptomatic primary breast cancer with mr imaging. *American Journal of Roentgenology*, 169(2):417–424, 1997.

[94] PS Myles and J Cui. Using the bland-altman method to measure agreement with repeated measures. *British Journal of Anaesthesia*,

99(3):309, 2007.

[95] Alessandro Neri, Daniele Marrelli, Tiziana Megha, Francesca Bettarini, Damiana Tacchini, Lorenzo De Franco, and Franco Roviello. Clinical significance of multifocal and multicentric breast cancers and choice of surgical treatment: a retrospective study on a series of 1158 cases. *BMC surgery*, 15(1):1–10, 2015.

[96] Calvin F Nodine, Claudia Mello-Thoms, Susan P Weinstein, Harold L Kundel, Emily F Conant, Rose E Heller-Savoy, Susan E Rowlings, and Julia A Birnbaum. Blinded review of retrospectively visible unreported breast cancers: An eye-position analysis 1. *Radiology*, 221(1):122–129, 2001.

[97] Mitra Noroozian, Lubomir Hadjiiski, Sahand Rahnama-Moghadam, Katherine A Klein, Deborah O Jeffries, Renee W Pinsky, Heang-Ping Chan, Paul L Carson, Mark A Helvie, and Marilyn A Roubidoux. Digital breast tomosynthesis is comparable to mammographic spot views for mass characterization. *Radiology*, 262(1):61–68, 2012.

[98] Jum C Nunnally. Psychometric theory. 1967.

[99] S Greenstein Orel, Mitchell D Schnall, Colleen M Powell, Mary G Hochman, Lawrence J Solin, Barbara L Fowble, Michael H Torosian, and Ernest F Rosato. Staging of suspected breast cancer: effect of mr imaging and mr-guided biopsy. *Radiology*, 196(1):115–122, 1995.

[100] Robert T Osteen and Lucy Hynds Karnell. The national cancer data base report on breast cancer. *Cancer*, 73(7), 1994.

[101] S Park, G Zhang, and K Myers. Comparison of channel methods and observer models for the task-based assessment of multi-projection imaging in the presence of structured anatomical noise. *IEEE Transactions on Medical Imaging*, 35(6):1431–1442, 2016.

[102] Subok Park, Aldo Badano, Brandon D Gallas, and Kyle J Myers. Incorporating human contrast sensitivity in model observers for detection tasks. *Medical Imaging, IEEE Transactions on*, 28(3):339–347, 2009.

[103] Subok Park, Harrison H Barrett, Eric Clarkson, Matthew A Kupinski, and Kyle J Myers. Channelized-ideal observer using laguerre-gauss channels in detection tasks involving non-gaussian distributed lumpy backgrounds and a gaussian signal. *JOSA A*, 24(12):B136–B150, 2007.

[104] Subok Park, Robert Jennings, Haimo Liu, Aldo Badano, and Kyle Myers. A statistical, task-based evaluation method for three-dimensional x-ray breast imaging systems using variable-background phantoms. *Medical Physics*, 37(12):6253–6270, 2010.

[105] Subok Park, Matthew A Kupinski, Eric Clarkson, and Harrison H Barrett. Ideal-observer performance under signal and background uncertainty. In *Biennial International Conference on Information Processing in Medical Imaging*, pages 342–353. Springer, 2003.

[106] Subok Park, Joel M Witten, and Kyle J Myers. Singular vectors of a linear imaging system as efficient channels for the bayesian ideal observer. *IEEE Transactions on Medical Imaging*, 28(5):657–668, 2009.

[107] Subok Park, George Z Zhang, Rongping Zeng, and Kyle J Myers. Comparing observer models and feature selection methods for a task-based statistical assessment of digital breast tomsynthesis in reconstruction space. In *SPIE Medical Imaging*, pages 90370M–90370M. International Society for Optics and Photonics, 2014.

[108] Theodore J Passe, David A Bluemke, and Stanley S Siegelman. Tumor angiogenesis: tutorial on implications for imaging. *Radiology*, 203(3):593–600, 1997.

[109] Heather R Peppard, Brandi E Nicholson, Carrie M Rochman, Judith K Merchant, Ray C Mayo III, and Jennifer A Harvey. Digital breast tomosynthesis in the diagnostic setting: indications and clinical applications. *Radiographics*, 35(4):975–990, 2015.

[110] Snehal M Pinto Pereira, Valerie A McCormack, Sue M Moss, and Isabel dos Santos Silva. The spatial distribution of radiodense breast tissue: a longitudinal study. *Breast Cancer Research*, 11(3):1–12, 2009.

[111] Etta D Pisano, Constantine A Gatsonis, Martin J Yaffe, R Edward Hendrick, Anna NA Tosteson, Dennis G Fryback, Lawrence W Bassett, Janet K Baum, Emily F Conant, Roberta A Jong, et al. American college of radiology imaging network digital mammographic imaging

screening trial: Objectives and methodology 1. *Radiology*, 236(2):404–412, 2005.

[112] Etta D Pisano, R Edward Hendrick, Martin J Yaffe, Janet K Baum, Suddhasatta Acharyya, Jean B Cormack, Lucy A Hanna, Emily F Conant, Laurie L Fajardo, Lawrence W Bassett, et al. Diagnostic accuracy of digital versus film mammography: exploratory analysis of selected population subgroups in dmist 1. *Radiology*, 246(2):376–383, 2008.

[113] Ljiljana Platiša, Bart Goossens, Ewout Vansteenkiste, Aldo Badano, and Wilfried Philips. Channelized hotelling observers for the detection of 2d signals in 3d simulated images. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 1781–1784. IEEE, 2009.

[114] Ljiljana Platiša, Bart Goossens, Ewout Vansteenkiste, Subok Park, Brandon D Gallas, Aldo Badano, and Wilfried Philips. Channelized hotelling observers for the assessment of volumetric imaging data sets. *JOSA A*, 28(6):1145–1163, 2011.

[115] Lucreţiu M Popescu. Model for the detection of signals in images with multiple suspicious locations. *Medical Physics*, 35(12):5565–5574, 2008.

[116] Lucreţiu M Popescu and Robert M Lewitt. Small nodule detectability evaluation using a generalized scan-statistic model. *Physics in Medicine and Biology*, 51(23):6225, 2006.

[117] Jacob Reinhold, Gezheng Wen, Joseph Y Lo, and Mia K Markey. Lesion detectability in stereoscopically viewed digital breast tomosynthesis projection images: a model observer study with anthropomorphic computational breast phantoms. In *SPIE Medical Imaging*, pages 101360W–101360W. International Society for Optics and Photonics, 2017.

[118] I Reiser, A Edwards, and RM Nishikawa. Validation of a power-law noise model for simulating small-scale breast tissue. *Physics in Medicine and Biology*, 58(17):6011, 2013.

[119] I Reiser and RM Nishikawa. Task-based assessment of breast tomosynthesis: Effect of acquisition parameters and quantum noisea). *Medical Physics*, 37(4):1591–1600, 2010.

[120] Baorui Ren, Andrew P Smith, and Zhenxue Jing. Local versus whole breast volumetric breast density assessments and implications. In *Breast Imaging*, pages 775–782. Springer, 2012.

[121] Paul Peter Rosen. *Rosen's breast pathology.* Lippincott Williams & Wilkins, 2001.

[122] Mark Ruschin, Pontus Timberg, Magnus Båth, Bengt Hemdal, Tony Svahn, Rob S Saunders, Ehsan Samei, Ingvar Andersson, Sören Mattsson, Dev P Chakraborty, et al. Dose dependence of mass and microcalcification detection in digital mammography: free response human observer studies. *Medical Physics*, 34(2):400–407, 2007.

[123] Murray B Sachs, Jacob Nachmias, and John G Robson. Spatial-frequency channels in human vision. *JOSA*, 61(9):1176–1186, 1971.

[124] Ehsan Samei, Robert S Saunders Jr, Jay A Baker, and David M Delong. Digital mammography: Effects of reduced radiation dose on diagnostic performance 1. *Radiology*, 243(2):396–404, 2007.

[125] Salim Samuel, Harold L Kundel, Calvin F Nodine, and Lawrence C Toto. Mechanism of satisfaction of search: eye position recordings in the reading of chest radiographs. *Radiology*, 194(3):895–902, 1995.

[126] Francesco Sardanelli, Gian M Giuseppetti, Pietro Panizza, Massimo Bazzocchi, Alfonso Fausto, Giovanni Simonetti, Vincenzo Lattanzio, and Alessandro Del Maschio. Sensitivity of mri versus mammography for detecting foci of multifocal, multicentric breast cancer in fatty and dense breasts using the whole-breast pathologic examination as a gold standard. *American Journal of Roentgenology*, 183(4):1149–1157, 2004.

[127] Ioannis Sechopoulos. A review of breast tomosynthesis. part i. the image acquisition process. *Medical Physics*, 40(1), 2013.

[128] Ioannis Sechopoulos. A review of breast tomosynthesis. part ii. image reconstruction, processing and analysis, and advanced applications. *Medical Physics*, 40(1), 2013.

[129] Ioannis Sechopoulos, Kristina Bliznakova, and Baowei Fei. Power spectrum analysis of the x-ray scatter signal in mammography and breast

tomosynthesis projections. *Medical Physics*, 40(10), 2013.

[130] Ioannis Sechopoulos and Caterina Ghetti. Optimization of the acquisition geometry in digital tomosynthesis of the breast. *Medical Physics*, 36(4):1199–1207, 2009.

[131] EA Sickles, CJ DOrsi, LW Bassett, American College of Radiology, et al. Acr bi-rads® mammography. *ACR BI-RADS Atlas, Breast Imaging Reporting and Data System*, pages 141–167, 2013.

[132] Robert L Siddon. Fast calculation of the exact radiological path for a three-dimensional ct array. *Medical Physics*, 12(2):252–255, 1985.

[133] Emil Y Sidky, Yuval Duchin, Ingrid Reiser, Christer Ullberg, and Xiaochuan Pan. Optimizing algorithm parameters based on a model observer detection task for image reconstruction in digital breast tomosynthesis. In *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2011 IEEE*, pages 4230–4232. IEEE, 2011.

[134] Emil Y Sidky, Xiaochuan Pan, Ingrid S Reiser, Robert M Nishikawa, Richard H Moore, and Daniel B Kopans. Enhanced imaging of microcalcifications in digital breast tomosynthesis through improved image-reconstruction algorithms. *Medical Physics*, 36(11):4920–4932, 2009.

[135] Luis Sisternes, Jovan G Brankov, Adam M Zysk, Robert A Schmidt, Robert M Nishikawa, and Miles N Wernick. A computational model to

generate simulated three-dimensional breast masses. *Medical Physics*, 42(2):1098–1118, 2015.

[136] Per Skaane, Andriy I Bandos, Ellen B Eben, Ingvild N Jebsen, Mona Krager, Unni Haakenaasen, Ulrika Ekseth, Mina Izadi, Solveig Hofvind, and Randi Gullien. Two-view digital breast tomosynthesis screening with synthetically reconstructed projection images: comparison with digital breast tomosynthesis with full-field digital mammographic images. *Radiology*, 271(3):655–663, 2014.

[137] Leslie H Sobin and Irvin D Fleming. Tnm classification of malignant tumors, (1997). *Cancer*, 80(9):1803–1804, 1997.

[138] Justin Solomon and Ehsan Samei. A generic framework to simulate realistic lung, liver and renal pathologies in ct imaging. *Physics in Medicine and Biology*, 59(21):6637, 2014.

[139] Gregory M Sturgeon, Nooshin Kiarashi, Joseph Y Lo, E Samei, and WP Segars. Finite-element modeling of compression and gravity on a population of breast phantoms for multimodality imaging simulation. *Medical Physics*, 43(5):2207–2217, 2016.

[140] TM Svahn, N Houssami, I Sechopoulos, and Sören Mattsson. Review of radiation dose estimates in digital breast tomosynthesis relative to those in two-view full-field digital mammography. *The Breast*, 24(2):93–99, 2015.

[141] Alberto Tagliafico, Davide Astengo, Francesca Cavagnetto, Raffaella Rosasco, Giuseppe Rescinito, Francesco Monetti, and Massimo Calabrese. One-to-one comparison between digital spot compression view and digital breast tomosynthesis. *European radiology*, 22(3):539–544, 2012.

[142] Hsin-Wu Tseng, Jiahua Fan, Matthew A Kupinski, Paavana Sainath, and Jiang Hsieh. Assessing image quality and dose reduction of a new x-ray computed tomography iterative reconstruction algorithm using model observers. *Medical Physics*, 41(7), 2014.

[143] Montpellier University Hospital. Value of tomosynthesis in breast lesion characterization and breast cancer staging.

[144] Marte Wasserman, Cindy Davis, Diana Edgar, and I. Monteiro. Diagnosis of US/MRI-occult asymptomatic multifocal invasive lobular carcinoma using breast tomosynthesis. *Applied Radiology*, 41(11):7–8, 2012.

[145] Lincoln J Webb, Ehsan Samei, Joseph Y Lo, Jay A Baker, Sujata V Ghate, Connie Kim, Mary Scott Soo, and Ruth Walsh. Comparative performance of multiview stereoscopic and mammographic display modalities for breast lesion detection. *Medical Physics*, 38(4):1972–1980, 2011.

[146] Tobias M Weissenbacher, Madeleine Zschage, Wolfgang Janni, Udo Jeschke, Thomas Dimpfl, Doris Mayr, Brigitte Rack, Christian Schindlbeck, Klaus

Friese, and Darius Dian. Multicentric and multifocal versus unifocal breast cancer: is the tumor-node-metastasis classification justified? *Breast Cancer Research and Treatment*, 122(1):27–34, 2010.

[147] Gezheng Wen, Tamara Miner Haygood, and Mia K Markey. A human observer study of multi-lesion detection in digital breast tomosynthesis. In *Medical Image Perception Society Meeting XVII*. Medical Image Perception Society Meeting, 2017.

[148] Gezheng Wen, Mia K Markey, Tamara Miner Haygood, and Subok Park. Digital breast tomosynthesis for multi-lesion detection in the presence of anatomical noise. *in preparation*, 2017.

[149] Gezheng Wen, Mia K Markey, Tamara Miner Haygood, and Subok Park. Model observer for assessing digital breast tomosynthesis for multi-lesion detection in the presence of anatomical noise. *in preparation*, 2017.

[150] Gezheng Wen, Mia K Markey, and Subok Park. Influence of local anatomical variations on detection of multifocal and multicentric breast cancer. *Medical Physics*, 43(6):3817–3817, 2016.

[151] Gezheng Wen, Mia K Markey, and Subok Park. Model observer design for detecting multiple abnormalities in anatomical background images. In *SPIE Medical Imaging*, page 97870S. International Society for Optics and Photonics, 2016.

[152] Gezheng Wen, Mia K Markey, and Subok Park. Model observer design for multi-signal detection in the presence of anatomical noise. *Physics in Medicine and Biology*, 62(4):1396, 2017.

[153] Gezheng Wen, Subok Park, and Mia K Markey. Digital breast tomosynthesis for detecting multifocal and multicentric breast cancer: influence of acquisition geometry on model observer performance in breast phantom images. In *SPIE Medical Imaging*, pages 101360V–101360V. International Society for Optics and Photonics, 2017.

[154] Joel M Witten, Subok Park, and Kyle J Myers. Partial least squares: a method to estimate efficient channels for the ideal observers. *IEEE Transactions on Medical Imaging*, 29(4):1050–1058, 2010.

[155] R Wolters, A Wöckel, W Janni, I Novopashenny, F Ebner, R Kreienberg, M Wischnewsky, L Schwentner, BRENDA Study Group, et al. Comparing the outcome between multicentric and multifocal breast cancer: what is the impact on survival, and is there a role for guideline-adherent adjuvant therapy? a retrospective multicenter cohort study of 8,935 patients. *Breast cancer research and treatment*, 142(3):579–590, 2013.

[156] Gang Wu, James G Mainprize, John M Boone, and Martin J Yaffe. Evaluation of scatter effects on image quality for breast tomosynthesis. *Medical Physics*, 36(10):4425–4432, 2009.

[157] Adam Wunderlich and Craig K Abbey. Utility as a rationale for choosing observer performance assessment paradigms for detection tasks in

medical imaging. *Medical Physics*, 40(11):111903, 2013.

[158] Stefano Young, Predrag R Bakic, Kyle J Myers, Robert J Jennings, and Subok Park. A virtual trial framework for quantifying the detectability of masses in breast tomosynthesis projection data. *Medical Physics*, 40(5):051914, 2013.

[159] Rongping Zeng, Aldo Badano, and Kyle J Myers. Optimization of digital breast tomosynthesis (dbt) acquisition parameters for human observers: effect of reconstruction algorithms. *Physics in Medicine and Biology*, 62(7):2598, 2017.

[160] Rongping Zeng, Subok Park, Predrag Bakic, and Kyle J Myers. Evaluating the sensitivity of the optimization of acquisition geometry to the choice of reconstruction algorithm in digital breast tomosynthesis through a simulation study. *Physics in Medicine and Biology*, 60(3):1259, 2015.

[161] George Z Zhang, Kyle J Myers, and Subok Park. Investigating the feasibility of using partial least squares as a method of extracting salient information for the evaluation of digital breast tomosynthesis. In *SPIE Medical Imaging*, pages 867311–867311. International Society for Optics and Photonics, 2013.

[162] Yani Zhang, Binh T Pham, and Miguel P Eckstein. The effect of nonlinear human visual system components on performance of a channelized

hotelling observer in structured backgrounds. *IEEE Transactions on Medical Imaging*, 25(10):1348–1362, 2006.

[163] Yani Zhang, Binh T Pham, and Miguel P Eckstein. Evaluation of internal noise methods for hotelling observer models. *Medical Physics*, 34(8):3312–3322, 2007.

[164] Yiheng Zhang, Heang-Ping Chan, Mitchell M Goodsitt, Andrea Schmitz, Jeffrey W Eberhard, and Bernhard EH Claus. Investigation of different pv distributions in digital breast tomosynthesis (dbt) mammography. In *International Workshop on Digital Mammography*, pages 593–600. Springer, 2008.

[165] Bo Zhao and Wei Zhao. Imaging performance of an amorphous selenium digital mammography detector in a breast tomosynthesis system. *Medical Physics*, 35(5):1978–1987, 2008.

[166] Bo Zhao and Wei Zhao. Three-dimensional linear system analysis for breast tomosynthesis. *Medical Physics*, 35(12):5219–5232, 2008.

[167] Mei-rong Zhou, Zhong-hua Tang, Jing Li, Jin-Hu Fan, Yi Pang, Hong-jian Yang, Shan Zheng, Jing-qiao Bai, Ning Lv, You-Lin Qiao, et al. Clinical and pathologic features of multifocal and multicentric breast cancer in chinese women: a retrospective cohort study. *Journal of breast cancer*, 16(1):77–83, 2013.

[168] Margarita L Zuley, Ben Guo, Victor J Catullo, Denise M Chough, Amy E Kelly, Amy H Lu, Grace Y Rathfon, Marion Lee Spangler, Jules H Sumkin, Luisa P Wallace, et al. Comparison of two-dimensional synthesized mammograms versus original digital mammograms alone and in combination with tomosynthesis images. *Radiology*, 271(3):664–671, 2014.

Vita

Gezheng Wen was born in Jinhua, China. He received his Bachelor degree in Electronic and Communications Engineering with a minor in Finance from The University of Hong Kong (HKU) in 2011. He was awarded Chiap Hua Chengs Foundation Scholarships, The Centenary Prize, CMA and Donors Scholarship, The Deans Honor List, and EEE Entrance Scholarship for his academic achievements at HKU. Gezheng joined The University of Texas at Austin (UT Austin) in 2011 to pursuit his Ph.D. degree in Electrical and Computer Engineering under the supervision of Dr. Mia K. Markey. In 2014, he received his Master of Science degree in Electrical and Computer Engineering from UT Austin. Gezheng has been a recipient of Professional Development Award, The Center for Perceptual Systems Travel Grant, Gordon Research Conference-Image Science Scholarship, and Medical Image Perception Society Scholarship. During his studies at UT Austin, Gezheng was a research trainee in Diagnostic Radiology at The University of Texas MD Anderson Cancer Center. He worked as a CT Clinical Science Co-op at Philips Healthcare from September 2014 to May 2015. He was a Scholar-in-Residence research intern at The U.S. Food and Drug Administration from June 2015 to August 2015.

Permanent address: 288 Shida Street, Apt 11-303, Lize Garden
Jinhua, Zhejiang Province, China 321004

This dissertation was typeset with LaTeX[†] by the author.

[†] LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.